SOME NON-LINEAR PROBLEMS IN APPROXIMATION*

BY

DUNHAM JACKSON

1. **Introduction.** Much has been written on the least-square properties of developments in series of orthogonal functions. The approximating function, characterized by the requirement that it shall reduce the integral of the square of the error to a minimum, is chosen in each case from a linear family (for example, the family of all trigonometric sums of the nth order), containing the sum of any two of its members, or the product of any of them by a constant. The writer has considered the problem of least-square approximation, and more generally of approximation according to the criterion of least mth powers, in terms of the non-linear family of all functions $\phi(x)$ which satisfy the condition

$$|\phi(x_2) - \phi(x_1)| \leqq \lambda |x_2 - x_1|$$

for a given value of λ, and has discussed the convergence of the approximation as λ becomes infinite. Even in this case, the *average* of any two members of the family corresponding to a specified λ is contained in the family; and this fact enters into the proofs of uniqueness.†

In the problems of the present paper, an approximating function is to be chosen from a family which does not in general contain the average of two of its members. The cases taken as illustrative are those in which the approximating function is the square of a trigonometric sum of the nth order, or the square root of a (positive) trigonometric sum of the nth order, the function to be represented being itself positive. Questions of uniqueness, which would at any rate call for novel methods of treatment, are allowed to lapse. Even so, it is found possible to deal with the convergence of the approximations as the order of the sums becomes infinite. The method is essentially that which the writer has used repeatedly in connection with problems of approximation by the method of least mth powers. It involves, however, an extension of Bernstein's theorem, which will be obtained in the next section.

2. **Extension of Bernstein's theorem.** The familiar statement of Bernstein's theorem is that if $T_n(x)$ is a trigonometric sum of the nth order such

* Presented to the Society, December 28, 1926; received by the editors October 12, 1927.

† See D. Jackson, *On approximation by functions of given continuity*, these Transactions, vol. 25 (1923), pp. 449–458; p. 450.

that $|T_n(x)| \leqq L$ for all values of x, then $|T_n'(x)|$ can not exceed nL. The statement can be generalized as follows:

If $f(x)$ is a function of period 2π having a continuous first derivative subject everywhere to the condition $|f'(x)| \leqq \lambda$, and if $T_n(x)$ is a trigonometric sum of the nth order such that $|T_n(x) - f(x)| \leqq L$ for all values of x, then

$$|T_n'(x)| \leqq nL + C\lambda, \tag{1}$$

where C is an absolute constant. More definitely (though the preceding statement is sufficient for the applications)

$$|T_n'(x)| \leqq nL + 4\lambda.$$

This formulation reduces to the standard one if $f(x)$ is identically zero.

Since $f(x)$ satisfies a Lipschitz condition with coefficient λ, there exists a trigonometric sum $t_n(x)$, of the nth order, such that

$$|t_n(x) - f(x)| \leqq \frac{K\lambda}{n}, \tag{2}$$

where K is an absolute constant.* Such a sum is defined by the formula

$$t_n(x) = I_m(x) = \frac{1}{2}h_m \int_{-\pi}^{\pi} f(v)\Phi_m(v - x)dv, \tag{3}$$

where

$$\Phi_m(v) = \frac{\sin^4(mv/2)}{m^4 \sin^4(v/2)}, \quad \frac{2}{h_m} = \int_{-\pi}^{\pi} \Phi_m(v)dv = \int_{-\pi}^{\pi} \Phi_m(v - x)dv,$$

the value of the last expression being only apparently and not actually dependent on x. The conditions for differentiating under the sign of integration are satisfied in (3), so that

$$\begin{aligned} t_n'(x) &= \frac{1}{2}h_m \int_{-\pi}^{\pi} f(v)\frac{\partial}{\partial x}\Phi_m(v - x)dv \\ &= -\frac{1}{2}h_m \int_{-\pi}^{\pi} f(v)\frac{\partial}{\partial v}\Phi_m(v - x)dv \\ &= \frac{1}{2}h_m \int_{-\pi}^{\pi} f'(v)\Phi_m(v - x)dv, \end{aligned}$$

* For the theorem, as well as for the details of proof referred to in the text, see D. Jackson, *On approximation by trigonometric sums and polynomials*, these Transactions, vol. 13 (1912), pp. 491–515; pp. 492–494.

the last form resulting from an integration by parts, with attention to the periodicity of the functions involved. Hence

$$|\, t_n'(x)\,| \leqq \frac{1}{2} h_m \lambda \int_{-\pi}^{\pi} \Phi_m(v - x)\,dv = \lambda. \tag{4}$$

Let $\tau_n(x) = T_n(x) - t_n(x)$. Then

$$|\, \tau_n(x)\,| = |\, [T_n(x) - f(x)] - [t_n(x) - f(x)]\,| \leqq L + \frac{K\lambda}{n}.$$

As $\tau_n(x)$ is a trigonometric sum, Bernstein's theorem is applicable, to the effect that

$$|\, \tau_n'(x)\,| \leqq nL + K\lambda. \tag{5}$$

As $T_n'(x) = \tau_n'(x) + t_n'(x)$, it follows from (4) and (5) that

$$|\, T_n'(x)\,| \leqq nL + (K+1)\lambda = nL + C\lambda, \quad C = K + 1.$$

Since (2) is true* with $K = 3$, the conclusion (1) holds for $C = 4$.

An essential point for the applications is that C is independent of n. As a constant may be subtracted from $f(x)$ and from $T_n(x)$ without affecting the essential conditions of the problem, it may be assumed without loss of generality that $f(x)$ vanishes somewhere in a period, and then it is obvious (with the use of Bernstein's theorem) that

$$|\, f(x)\,| \leqq \pi\lambda, \quad |\, T_n(x)\,| \leqq L + \pi\lambda, \quad |\, T_n'(x)\,| \leqq n(L + \pi\lambda).$$

But this comparatively trivial observation would not serve the purpose.

3. **Approximation by squares of trigonometric sums.** Let $f(x)$ be a given function of period 2π, which is positive everywhere and which has everywhere a continuous first derivative, and let λ be the maximum of $|f'(x)|$. Among all trigonometric sums of given order n, let $T_n(x)$ be one for which the integral

$$\int_{-\pi}^{\pi} \{f(x) - [T_n(x)]^2\}^2 dx$$

has the smallest possible value. It can be inferred readily from well known theorems that the greatest lower bound of this expression is a minimum which is actually attained. For if an upper bound is assigned to the value of the integral, the coefficients in $[T_n(x)]^2$, considered as a trigonometric sum of order $2n$, must belong to a bounded domain; the maximum of $[T_n(x)]^2$, and hence that of $|T_n(x)|$, is thereby restricted; and so the coefficients that

* These Transactions, vol. 13, loc. cit., Theorem VI, p. 510.

come into consideration for $T_n(x)$ belong to a bounded region, which may be taken as closed, so that the fundamental theorem on the existence of a minimum is applicable. On the other hand, the average of the squares of two trigonometric sums is not in general the square of a trigonometric sum,* and because of this circumstance the proof commonly given for the uniqueness of the minimum in similar cases breaks down. The determination of $T_n(x)$ is manifestly not unique, since $[-T_n(x)]$ would serve the same purpose; the question of the uniqueness of $[T_n(x)]^2$ will be left in abeyance. It will be supposed that a particular $T_n(x)$ minimizing the integral is designated for each value of n, and it will be shown that $[T_n(x)]^2$ *converges uniformly toward* $f(x)$ *as* n *becomes infinite.*

With the understanding, then, that $T_n(x)$ is a sum which minimizes the integral for a specified value of n, let

$$\gamma_n = \int_{-\pi}^{\pi} \{f(x) - [T_n(x)]^2\}^2 dx.$$

Let

$$R_n(x) = f(x) - [T_n(x)]^2,$$

let μ_n be the maximum of $|R_n(x)|$, and let x_0 be a value of x such that

$$|R_n(x_0)| = \mu_n.$$

As $[T_n(x)]^2$ is a trigonometric sum of order $2n$, and as $|f'(x)| \leqq \lambda$, it follows from the preceding section that

$$\left|\frac{d}{dx}[T_n(x)]^2\right| \leqq 2n\mu_n + C\lambda, \quad |R_n'(x)| \leqq 2n\mu_n + (C+1)\lambda.$$

Let it be supposed temporarily that $(C+1)\lambda \leqq n\mu_n$, the contrary case being reserved for later consideration. Then

$$|R_n'(x)| \leqq 3n\mu_n, \quad |R_n(x) - R_n(x_0)| \leqq 3n\mu_n |x - x_0|,$$

and for $|x - x_0| \leqq 1/(6n)$,

$$|R_n(x) - R_n(x_0)| \leqq \tfrac{1}{2}\mu_n, \qquad |R_n(x)| \geqq \tfrac{1}{2}\mu_n.$$

Since the last relation holds throughout an interval of length $1/(3n)$ at least,

$$\gamma_n \geqq (\tfrac{1}{2}\mu_n)^2/(3n).$$

On the other hand, since $f(x)$ has a positive minimum and possesses a continuous derivative, $[f(x)]^{1/2}$ also has a continuous derivative, and so

* E.g., the average of $2\sin^2 x$ and $4\cos^2 x$ is $1+\cos^2 x$, and it is readily seen that no trigonometric sum can have the last expression for its square.

satisfies a Lipschitz condition, with some coefficient which may be denoted by λ'. Therefore a trigonometric sum $t_n(x)$ of the nth order can be constructed so that

$$| [f(x)]^{1/2} - t_n(x) | \leqq K\lambda'/n,$$

where K is the absolute constant already cited. If such sums are formed for all positive integral values of n, they are uniformly bounded; the expressions

$$| [f(x)]^{1/2} + t_n(x) |$$

have an upper bound A which is independent of n, and

$$| f(x) - [t_n(x)]^2 | \leqq AK\lambda'/n.$$

By reason of the minimizing property of $T_n(x)$,

$$\gamma_n \leqq \int_{-\pi}^{\pi} \{f(x) - [t_n(x)]^2\}^2 dx \leqq 2\pi(AK\lambda'/n)^2.$$

If $2\pi(AK\lambda')^2 = B$, then,

$$(\tfrac{1}{2}\mu_n)^2/(3n) \leqq \gamma_n \leqq B/n^2, \qquad \mu_n \leqq 2(3B)^{1/2}/n^{1/2}.$$

These relations have been deduced on the hypothesis that $(C+1)\lambda \leqq n\mu_n$. To deny this hypothesis, however, is to suppose directly that $\mu_n < (C+1)\lambda/n$. If D is the larger of the numbers $2(3B)^{1/2}$, $(C+1)\lambda$, it is certain in any case that

$$\mu_n \leqq D/n^{1/2},$$

which means that $\lim_{n=\infty} \mu_n = 0$, and the convergence is proved.

4. **Approximation by square roots of trigonometric sums.** Let $f(x)$ be subject to the same hypotheses as before. Let $T_n(x)$ be characterized this time, however, among all non-negative trigonometric sums of the nth order, by the requirement that the integral

$$\int_{-\pi}^{\pi} \{f(x) - [T_n(x)]^{1/2}\}^2 dx$$

shall be a minimum. In the preceding section, the approximating function belonged to a certain restricted class of trigonometric sums; here it is not in general a trigonometric sum at all.

Informally stated, the argument for the existence of a minimum is as follows. Consider the integral

$$\int_{-\pi}^{\pi} \{f(x) - [t_n(x)]^{1/2}\}^2 dx$$

formed with an arbitrary trigonometric sum $t_n(x)$ of the nth order. If any coefficient in $t_n(x)$ were very large, the value of the expression in braces would be large at some point. The magnitude of its derivative being restricted (somewhat indirectly, but in fairly obvious fashion) by Bernstein's theorem, the expression itself would necessarily remain large over an appreciable interval, and the integral would be large. The approach of the integral to its greatest lower bound can take place therefore only in a restricted domain for the coefficients, where the fundamental theorem on the existence of a minimum is once more in force.

The method that first comes to mind in connection with an attempt to establish the uniqueness of the minimum is blocked again, by the fact this time that the average of the square roots of two trigonometric sums is not in general the square root of a trigonometric sum.* The question of uniqueness will accordingly be dismissed from further consideration, and it will be supposed merely that $T_n(x)$ denotes for each positive integral n a particular sum of the nth order for which the integral has its minimum value, whether it be the only sum having this property or not.

With this understanding, let

$$\gamma_n = \int_{-\pi}^{\pi} \{f(x) - [T_n(x)]^{1/2}\}^2 dx.$$

The problem is to show that $[T_n(x)]^{1/2}$ converges uniformly toward $f(x)$. Because of the fact that $[T_n(x)]^{1/2}$ is not a trigonometric sum, several preliminary steps are needed before the lemma of §2 can be applied. Let λ' be the maximum of $|2f(x)f'(x)|$. Then $[f(x)]^2$ satisfies a Lipschitz condition with coefficient λ', and there exists a trigonometric sum $t_n(x)$, of the nth order, such that

$$|[f(x)]^2 - t_n(x)| \leqq K\lambda'/n. \tag{6}$$

As $[f(x)]^2$ has a positive minimum, it is clear from (6) itself that $t_n(x)$ is everywhere positive, at least for values of n from a certain point on; if it is constructed according to the procedure used in establishing the general theorem to which reference is made (cf. §2) it is in fact never less than the minimum of $[f(x)]^2$. Let a be the minimum of $f(x)$, a positive number, by hypothesis. Then

$$f(x) + [t_n(x)]^{1/2} \geqq a,$$

* For example, the average of $2(2^{1/2}+\sin x)^{1/2}$ and $2(2^{1/2}-\sin x)^{1/2}$ is the square root of
$$2^{3/2}+2(2-\sin^2 x)^{1/2} = 2^{3/2}+2(1+\cos^2 x)^{1/2},$$
and $(1+\cos^2 x)^{1/2}$, as already remarked, is not a trigonometric sum.

the second term representing naturally the positive square root, and

$$|f(x) - [t_n(x)]^{1/2}| \leqq K\lambda'/(an).$$

Consequently

$$\gamma_n \leqq \int_{-\pi}^{\pi} \{f(x) - [t_n(x)]^{1/2}\}^2 dx \leqq 2\pi(K\lambda')^2/(an)^2. \tag{7}$$

Let M_n be the maximum of $[T_n(x)]^{1/2}$, and let M be the maximum of $f(x)$. In consideration of the alternatives $M_n < 2M$, $M_n \geqq 2M$, let it be supposed for the sake of argument that the latter relation is fulfilled. Let x_1 be a value of x for which $[T_n(x_1)]^{1/2} = M_n$. If $[T_n(x)]^{1/2}$ is not everywhere greater than $\frac{3}{4}M_n$, let x_2 be the value of x nearest to x_1 for which $[T_n(x_2)]^{1/2} = \frac{3}{4}M_n$, and let $|x_2 - x_1| = \delta$. Then

$$\left|\frac{d}{dx}[T_n(x)]^{1/2}\right| = \frac{|T_n'(x)|}{2[T_n(x)]^{1/2}} \leqq \frac{2|T_n'(x)|}{3M_n} \leqq \frac{|T_n'(x)|}{M_n},$$

for $|x - x_1| \leqq \delta$, since $[T_n(x)]^{1/2} \geqq \frac{3}{4}M_n$ for the values of x in question. But $|T_n(x)| \leqq M_n^2$, and hence $|T_n'(x)| \leqq nM_n^2$, by Bernstein's theorem. So

$$\left|\frac{d}{dx}[T_n(x)]^{1/2}\right| \leqq nM_n$$

for $|x - x_1| \leqq \delta$. This implies further that

$$\tfrac{1}{4}M_n = [T_n(x_1)]^{1/2} - [T_n(x_2)]^{1/2} \leqq nM_n|x_2 - x_1| = nM_n\delta,$$

$$\delta \geqq 1/(4n).$$

But as long as $|x - x_1| \leqq \delta$,

$$|f(x) - [T_n(x)]^{1/2}| \geqq \tfrac{3}{4}M_n - M \geqq \tfrac{1}{4}M_n, \tag{8}$$

since it is supposed for the time being that $M \leqq \frac{1}{3}M_n$, and consequently

$$\gamma_n \geqq 2\delta(\tfrac{1}{4}M_n)^2 \geqq 2\delta(\tfrac{1}{2}M)^2 \geqq M^2/(8n),$$

which is in contradiction with (7) for all values of n from a certain point on. To go back to one of the alternatives temporarily rejected, if $[T_n(x)]^{1/2} > \frac{3}{4}M_n$ everywhere, while M_n is still $\geqq 2M$, then (8) holds everywhere, and

$$\gamma_n \geqq 2\pi(\tfrac{1}{4}M_n)^2 \geqq 2\pi(\tfrac{1}{2}M)^2 = M^2\pi/2,$$

which again contradicts (7) from a certain point on. So the hypothesis that $M_n \geqq 2M$ can not be sustained for more than a finite number of values of n, which means that *the sums $T_n(x)$ are uniformly bounded.*

Let M', independent of n, be a common upper bound for the expressions $f(x)+[T_n(x)]^{1/2}$. Let

$$R_n(x) = f(x) - [T_n(x)]^{1/2},$$

and let μ_n be the maximum of $|R_n(x)|$. Then

$$|[f(x)]^2 - T_n(x)| \leqq M'\mu_n.$$

The maximum of the absolute value of the derivative of $[f(x)]^2$ has already been denoted by λ'. So the lemma of §2 yields the information that

$$|T_n'(x)| \leqq nM'\mu_n + C\lambda'. \tag{9}$$

The minimum of $f(x)$ is $a>0$. It is certain that $[T_n(x)]^{1/2}$ is not everywhere less than a, for any specified value of n, since the substitution of the constant a^2 for $T_n(x)$ would then give the integral a smaller value than the alleged minimum γ_n. If $[T_n(x)]^{1/2}$ has a minimum less than a, it must take on the value a and all values between a and the minimum. Suppose for the moment that the minimum is less than $\frac{1}{2}a$. Let y_1 be a point where $[T_n(x)]^{1/2} = \frac{3}{4}a$, and y_2 a point at which $[T_n(x)]^{1/2} = \frac{1}{2}a$, y_1 and y_2 being adjacent points of their respective categories, so that neither of the values $\frac{1}{2}a$, $\frac{3}{4}a$ is taken on anywhere between them. Since $[T_n(x)]^{1/2}$ is uniformly bounded, μ_n is bounded, and it follows from (9) that $|T_n'(x)|$ has an upper bound of the order of magnitude of n. Between y_1 and y_2, where $[T_n(x)]^{1/2}$ is never less than $\frac{1}{2}a$, the derivative of $[T_n(x)]^{1/2}$ is likewise less than a constant multiple of n in absolute value, say

$$\left|\frac{d}{dx}[T_n(x)]^{1/2}\right| < bn.$$

Hence

$$\frac{1}{4}a = [T_n(y_1)]^{1/2} - [T_n(y_2)]^{1/2} < bn|y_1 - y_2|,$$

$$|y_1 - y_2| > a/(4bn).$$

But $f(x)-[T_n(x)]^{1/2} \geqq \frac{1}{4}a$ throughout this interval, since the terms of the difference are respectively not less than a and not greater than $\frac{3}{4}a$. Consequently

$$\gamma_n \geqq a^3/(64bn),$$

which contradicts (7) as soon as n is sufficiently large. The contradiction arises from the hypothesis that the minimum of $[T_n(x)]^{1/2}$ is less than $\frac{1}{2}a$. Apart from a finite number of values of n, which can be left out of account without affecting the question of convergence, it must be that

$$[T_n(x)]^{1/2} \geqq \tfrac{1}{2}a \tag{10}$$

everywhere. It will be understood henceforth that this is the case.

Taken in conjunction with (10), the relation (9) signifies that

$$\left|\frac{d}{dx}[T_n(x)]^{1/2}\right| \leqq (nM'\mu_n + C\lambda')/a,$$

and if λ denotes the maximum of $|f'(x)|$, as before,

$$|R_n'(x)| \leqq Gn\mu_n + H,$$

where $G = M'/a$, $H = (C\lambda'/a) + \lambda$, both quantities being independent of n.

The rest of the proof runs along familiar lines. If $H > Gn\mu_n$, for a specified value of n, then

$$\mu_n < H/(Gn). \tag{11}$$

If $H \leqq Gn\mu_n$,

$$|R_n'(x)| \leqq 2Gn\mu_n,$$

and $|R_n(x)|$, attaining the value μ_n at some point, remains greater than or equal to $\frac{1}{2}\mu_n$ throughout an interval of length at least $1/(2Gn)$. This implies that $\gamma_n \geqq \mu_n^2/(8Gn)$. But it has been seen in (7) that $\gamma_n \leqq k_1/n^2$, if k_1 is used to denote the quantity $2\pi(K\lambda')^2/a^2$, independent of n. So, under the present hypothesis with regard to the relative magnitudes of H and $Gn\mu_n$,

$$\mu_n \leqq (8Gk_1/n)^{1/2}. \tag{12}$$

For every value of n, possibly with a finite number of exceptions, as previously noted, μ_n is subject to one or the other of the relations (11), (12), and certainly approaches zero as n becomes infinite. *This means that* $[T_n(x)]^{1/2}$ *converges uniformly toward* $f(x)$.

THE UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.

SECOND-ORDER DIFFERENTIAL SYSTEMS WITH INTEGRAL AND k-POINT BOUNDARY CONDITIONS*

BY

WILLIAM M. WHYBURN

Tamarkin† and C. E. Wilder‡ have considered the expansion problem for differential systems where the boundary conditions are of the integral and k-point types. Von Mises§ attacked the problem of proving existence and oscillation theorems for a special second-order equation with two integral conditions. However, von Mises did not carry his investigation to the point of establishing definite existence and oscillation theorems for his system. It is the purpose of the present paper to develop properties of second-order differential systems whose coefficients are Lebesgue summable functions and whose boundary conditions either contain definite integrals or apply to more than two points of the interval of definition.

1. Summable functions

Let $F(x)$ be any function that is defined on the finite interval $X: a \leqq x \leqq b$, and has the property that the Lebesgue integral $\int_a^b F(x)\,dx$ exists.

Let the interval X be divided into n subintervals $I_{in} = (x_{in}, x_{i+1,n})$ of length Δ_{in} by means of the $n+1$ distinct points $a \equiv x_{0n}, x_{1n}, \cdots, x_{nn} \equiv b$. Let $\max \Delta(n)$ represent the greatest Δ_{in} for each n. Then the sets of subdivision I_{in} represent a *fine* subdivision of X if $\lim_{n\to\infty} \max \Delta(n) = 0$. For a given set of subintervals I_{in} of X, $h_n(x)$ is a horizontal function of index n on X, if $h_n(x)$ has a constant value h_{in} for each value of x interior to I_{in}. The sum $\sum_{i=0,\cdots,n-1} h_{in}\Delta_{in}$ is called‖ the *area* under $h_n(x)$ on X.

THEOREM I. *There exists a sequence of horizontal functions $H_n(x)$ that is defined for every method of fine subdivision and has the following properties:*

(1) *For every method of fine subdivision of X, $\lim_{n\to\infty} H_n(x) = F(x)$ almost everywhere on X.*

---

* Presented to the Society, February 26, 1927, and April 15, 1927; received by the editors in May, 1927.

† *On some General Problems of the Theory*, etc., Petrograd, 1917.

‡ These Transactions, vol. 18 (1917), pp. 415–442.

§ *Festschrift Heinrich Weber*, Berlin, 1912, pp. 252–282.

‖ Cf. Ettlinger, American Journal of Mathematics, vol. 48 (1926), pp. 216–222.

(2) *If ϵ is a given arbitrary positive constant, there exist positive numbers M_ϵ and N_ϵ such that, if F_n denotes the set of points for which $|H_n(x)|>M_\epsilon$, then for all $n>N_\epsilon$, $\sum_{(F_n)}|H_{in}|\Delta_{in}<\epsilon$.*

(3) *If A_n denotes the area under $H_n(x)$ for each n, then the limit as n increases indefinitely of A_n exists independent of the method of fine subdivision and this limit is the Lebesgue integral of $F(x)$ taken over X.*

A sequence of horizontal functions that is readily seen to have these properties is defined by

$$H_{in}=\left[\int_{x_{in}}^{x_{i+1,n}}F(x)dx\right]/\Delta_{in}\qquad(i=1,2,\cdots,n).$$

THEOREM II. *If $h_n(x)$ is any sequence of horizontal functions that is defined for all methods of fine subdivision of X and has properties* (1) *and* (2) *of Theorem* I, *then $h_n(x)$ also has property* (3) *of that theorem.*

Theorem II follows from an application of the Duhamel-Moore theorem or the equivalent theorem given by Ettlinger.* The details of this proof are omitted here as the reader will have no difficulty in supplying them.

2. DIFFERENTIAL EQUATIONS WITH SUMMABLE COEFFICIENTS

Consider the pair of differential equations

$$\frac{dy}{dx}=K(x)z,\quad\frac{dz}{dx}=G(x)y,\tag{2.1}$$

where $K(x)>0$, $G(x)<0$, K and G are summable on X and have finite values at a and b.

THEOREM III. *If G/K is a non-increasing function of x on X and if c and $d(c<d)$ are consecutive† zeros of $y(x)$ on X, then $|z(c)|\leqq|z(d)|$.*

Multiply the second equation of (2.1) by z to get $zz'=Gyy'/K$ almost everywhere on X. Since $z(x)$ is absolutely continuous‡ on X, $\int_c^d zz'dt$ exists and hence $I=\int_c^d yy'G/Kdt$ exists and $I=\int_c^d zz'dt=(1/2)[z^2(d)-z^2(c)]$. Since G/K is monotonic and finite valued at a and b, it is bounded on X and from this fact, together with the summability of K and G, follows the summability of G/K. Let the points $x_{0n}=c$, x_{1n}, $\cdots$, $x_{nn}=d$ define a *fine*

* Loc. cit., p. 216.

† A proof that y and z have only a finite number of zeros on X is given by Ettlinger, Proceedings of the National Academy of Sciences, vol. 12 (1926), p. 540.

‡ See Carathéodory, *Vorlesungen über reele Funktionen*, Berlin, 1918, p. 678.

subdivision of $\overline{X}$: $c \leqq x \leqq d$ and let $H_{in} = [\int_{x_{in}}^{x_{i+1,n}} (G/K)dt]/\Delta_{in}$. The function $H_n(x)$ is bounded on $\overline{X}$ and $\lim_{n\to\infty} H_n(x) = G(x)/K(x)$ almost everywhere on this interval.

Let $H^*(x)$ be defined† by $H_{in}^* = [y_{in}\int_{x_{in}}^{x_{i+1,n}} y'(t)dt]/\Delta_{in}$. The function $H^*(x)$ has all three of the properties of Theorem I with respect to the summable function yy'. If we define a new sequence of horizontal functions $h_n(x)$ by letting $h_{in} = H_{in}H_{in}^*$, we see that $h_n(x)$ has properties (1) and (2) of Theorem I with respect to the summable function $yy'G/K$. It follows from Theorem II that

$$I = \lim_{n\to\infty} \sum_{i=0}^{n-1} h_{in}\Delta_{in}.$$

Now

$$\Delta[H_{in}y_{in}^2] = H_{in}[\Delta y_{in}]^2 + 2H_{in}y_{in}\,\Delta y_{in} + y_{i+1,n}^2\,\Delta H_{in}$$

and

$$\sum_{i=0}^{n-1}\Delta[H_{in}y_{in}^2] = \sum_{i=0}^{n-1} H_{in}[\Delta y_{in}]^2 + 2\sum_{i=0}^{n-1} H_{in}y_{in}\Delta y_{in} + \sum_{i=0}^{n-1} y_{i+1,n}^2\,\Delta H_{in}.$$

Since $y_{0n} = y(c) = y_{nn} = y(d) = 0$, the left hand side of the foregoing equation is zero and $2\sum_{i=0,\cdots,n-1} H_{in}y_{in}\Delta y_{in} = S_n + s_n$, where $S_n = -\sum_{i=0,\cdots,n-1} H_{in}\cdot[\Delta y_{in}]^2$ and $s_n = -\sum_{i=0,\cdots,n-1} \Delta H_{in}y_{i+1,n}^2$. If we let $\bar{h}_{in} = \Delta y_{in}H_{in}\Delta y_{in}/\Delta_{in}$, we note immediately that $\bar{h}_n(x)$ has property (2) and $\lim_{n\to\infty} \bar{h}_n(x) = 0$ almost everywhere on $\overline{X}$. It follows from Theorem II that $\lim_{n\to\infty} S_n = 0$. Hence

$$2I = 2\lim_{n\to\infty} \sum_{i=0}^{n-1} h_{in}\Delta_{in} = 2\lim_{n\to\infty} \sum_{i=0}^{n-1} H_{in}y_{in}\,\Delta y_{in} = 2\lim_{n\to\infty} s_n.$$

Since $\Delta H_{in} \leqq 0$, it follows that $s_n \geqq 0$, and $z^2(d) - z^2(c) = 2I \geqq 0$, or $|z(d)| \geqq |z(c)|$. This completes the proof of Theorem III.‡

The foregoing argument proves that $\lim_{n\to\infty} s_n = \int_c^d (yy'G/K)dt$. An examination of s_n shows that the above limit is the Stieltjes integral $\int_c^d y^2 \cdot d(G/K)$. Hence Stieltjes $\int_c^d y^2 d(G/K)$ = Lebesgue $\int_c^d (yy'G/K)dt$.

COROLLARY 1. *If G/K is a non-decreasing function of x on X, then* $|z(d)| \leqq |z(c)|$.

COROLLARY 2. *If G/K actually decreases over some sub-interval of $\overline{X}$, then*

$$|z(d)| > |z(c)|.$$

---

† $y_{in} = y(x_{in})$ and $\Delta y_{in} = y_{i+1,n} - y_{in}$. In general, $\Delta f_{in} = f_{i+1,n} - f_{in}$.

‡ If we add the requirement that $(G/K)'$ exist almost everywhere and be summable on X, this theorem follows from an application of partial integration.

COROLLARY 3. *If $y(c)\neq 0$ and d is the first zero of $y(x)$ such that $d>c$ then under the hypotheses of Theorem* III, $|z(c)|<|z(d)|$.

An examination of the proof for Theorem III shows that $I\geqq -(G/K)\cdot[y(c)]^2$ and this corollary follows.

The symmetry that exists between y and z yields the following:

THEOREM IV. *If K/G is a non-increasing function of x on X and if e and f $(e<f)$ are consecutive zeros of $z(x)$ on X, then $|y(e)|\leqq|y(f)|$.*

Theorem IV also yields corollaries analogous to these of Theorem III.

3. DIFFERENTIAL SYSTEMS WITH INTEGRAL BOUNDARY CONDITIONS

Consider the differential system

$$\frac{dy}{dx}=K(x,\lambda)z,\qquad \frac{dz}{dx}=G(x,\lambda)y\,; \tag{3.1}$$

$$\alpha(\lambda)z(a,\lambda)-\beta(\lambda)y(a,\lambda)=0,\quad F(b,\lambda)=0\,; \tag{3.2}$$

where $F(x,\lambda)=\int_a^x A(t,\lambda)y(t,\lambda)dt$, K, G and A are continuous functions of λ on L: $L_1<\lambda<L_2$ for each fixed x on X, summable functions of x on X for each fixed λ on L, and bounded numerically for all values of x and λ on XL by a function $M(x)$ that is summable on X. The functions $\alpha(\lambda)$ and $\beta(\lambda)$ are continuous in λ on L. We suppose that $K(x,\lambda)$ is positive on XL and that the coefficients K, G, α, and β satisfy conditions that are sufficient to insure the validity of existence and oscillation theorems* for the system (3.1), (3.3), where

$$\alpha(\lambda)z(a,\lambda)-\beta(\lambda)y(a,\lambda)=0,\quad y(b,\lambda)=0. \tag{3.3}$$

Case I. $\alpha(\lambda)\neq 0$ on L.

Carathéodory's existence theorem† insures the existence of a pair of functions, $y(x,\lambda)$ and $z(x,\lambda)$, that are continuous in both variables on XL, absolutely continuous in x for each fixed λ on L, satisfy (3.1) almost everywhere on X, and satisfy the conditions $y(a,\lambda)=\alpha(\lambda)$, $z(a,\lambda)=\beta(\lambda)$ for every λ on L. Consider this pair of functions.

THEOREM V. *If for a fixed value of λ on L, G/K is negative and a non-increasing function of x on X, and A/G is positive and a non-decreasing function of x on X, then the zeros of $F(x)$ and the zeros of $y(x)$ separate each other on X.*

---

* See Bôcher, *Leçons sur les Méthodes de Sturm*, Paris, 1917, p. 66.

† Loc. cit., p. 678.

Since the zeros of $y(x)$ on X are simple and finite in number, we will let $x_1<x_2<\cdots<x_n$ be these zeros. Let $I_0=\int_a^{x_1} Ay\,dt$ and $I_i=\int_{x_i}^{x_{i+1}} Ay\,dt$. Clearly $I_jI_{j+1}<0$ for $j=0,1,\cdots,n-1$. Hence if we show that $|I_0|<|I_1| \leqq |I_2|\leqq|I_3|\leqq\cdots\leqq|I_i|\leqq\cdots\leqq|I_{n-1}|$, it will follow that $F(x_i)\cdot F(x_{i+1})<0$, and since $F(x)$ is a continuous, increasing (or decreasing) function of x on $x_i\leqq x\leqq x_{i+1}$, we can conclude that $F(x)$ has one and only one zero on the interval $x_i\leqq x\leqq x_{i+1}$. To show that $|I_i|\leqq|I_{i+1}|$, $i>0$, we consider two cases. In the first case let I_i be positive. We wish to show that $I_i\leqq -I_{i+1}$. Replace y in the integrands by its value from (3.1). Since $Ay=Az'/G$ almost everywhere on X, it follows that the integrals are not changed by this substitution. From the monotonic character of A/G,

$$\int_{x_i}^{x_{i+1}}(Az'/G)dt\leqq\frac{A(x_{i+1})}{G(x_{i+1})}\int_{x_i}^{x_{i+1}}z'dt\,;$$

$$-\int_{x_{i+1}}^{x_{i+2}}(Az'/G)dt\geqq-\frac{A(x_{i+1})}{G(x_{i+1})}\int_{x_{i+1}}^{x_{i+2}}z'dt,$$

and hence we will prove a stronger inequality if we show that

$$-\frac{A(x_{i+1})}{G(x_{i+1})}\int_{x_{i+1}}^{x_{i+2}}z'dt\geqq\frac{A(x_{i+1})}{G(x_{i+1})}\int_{x_i}^{x_{i+1}}z'dt,$$

or $-z(x_{i+2})+z(x_{i+1})\geqq z(x_{i+1})-z(x_i)$, which is the same as $-z(x_{i+2})\geqq -z(x_i)$. Since $z(x)$ vanishes once on $x_{i+1}<x<x_{i+2}$ (its zeros separate those of y) and z' is negative wherever it exists on this interval, we have $z(x_{i+2})<0$. A similar argument gives $z(x_i)<0$. An application of Theorem III yields $|z(x_{i+2})|\geqq|z(x_{i+1})|\geqq|z(x_i)|$ and hence $-z(x_{i+2})\geqq-z(x_i)$. This establishes our inequality for the case $I_i>0$. The case $I_i<0$ is treated in the same way to get $z(x_{i+2})\geqq z(x_i)$, where $z(x_{i+2})$ and $z(x_i)$ are both positive. The validity of this inequality is a consequence of Theorem III.

A similar argument to that given above reduces the proof that $|I_0|<|I_1|$ to an application of Corollary 3, Theorem III. This completes the proof of our separation theorem.

THEOREM VI. *If $\lambda_0<\lambda_1<\lambda_2<\cdots$ are the ordered characteristic numbers of the system* (3.1), (3.3) *and if the hypotheses of Theorem* V *are satisfied for every λ on $\lambda_0\leqq\lambda<L_2$, then there exists an infinite set of characteristic values.* $k_0, k_1, k_2, \cdots$, for the system* (3.1), (3.2) *such that $L_1<\lambda_0<k_0<\lambda_1<k_1<\cdots<L_2$.*

* k_i may be an aggregate of a finite or infinite number of values.

$y(x, \lambda)$ and $z(x, \lambda)$ have been chosen so that the condition $\alpha(\lambda)z(a, \lambda) - \beta(\lambda)y(a, \lambda) = 0$ is satisfied for every λ on L. An implicit function theorem that was stated and proved "im kleinen" by Ettlinger* and extended by Sturdivant† to hold "im grossen," insures that the zeros of $F(x, \lambda)$ are continuous functions of λ on L. We extend the definitions of the coefficients of our system to apply to the interval $\overline{X}$: $a \leqq x \leqq b'$ $(b' > b)$ by making them have the same values outside of X that they have at $x = b$. For any fixed λ that is greater than (or equal to) λ_0, the zeros of F and y separate each other on X. Let us fix our attention on the ith zero that $F(x, \lambda)$ has on $a < x \leqq b'$ and let λ increase from λ_i to λ_{i+1}. If $x_i(\lambda)$ denotes this zero, then $x_i(\lambda_i) > b$ while $x_i(\lambda_{i+1}) < b$. Since $x_i(\lambda)$ is a continuous function, there must be at least one value of λ between λ_i and λ_{i+1} for which $x_i(\lambda) = b$. Let k_i be the set of all values of λ for which $x_i(\lambda) = b$. It follows from the continuity of $x_i(\lambda)$ and the separation of the zeros of F and y, that $\lambda_i < k_i < \lambda_{i+1}$.

THEOREM VII. *If p_i is any value of λ that belongs to k_i, then, on $a < x < b$, $F(x, p_i)$ has exactly i zeros, $y(x, p_i)$ has exactly $i+1$ zeros, and $z(x, p_i)$ has either i, $i+1$, or $i+2$ zeros.*

Since $y(a, \lambda) \neq 0$, it follows that no zeros of y or of F are lost from X. Since one and only one zero of y enters the X interval every time λ passes a characteristic value of the system (3.1), (3.3) a direct count reveals the validity of this theorem for $y(x, p_i)$. The remainder of the theorem follows from the separation of the zeros of $F(x, p_i)$ and $z(x, p_i)$, respectively, by the zeros of $y(x, p_i)$.

Case II. $\alpha(\lambda) \equiv 0$ and for every λ on L there exists a neighborhood of $x = a$ that is a subset of X, throughout which one of the quantities A/G, K/G actually increases as x increases.

The treatment of this case is the same as that given for Case I. The hypothesis that either A/G or K/G actually increase over some sub-interval of X that has a for one end point enables us to prove the same separation, existence, and oscillation theorems that were proved in Case I.

Case III. $\alpha(\lambda) \equiv 0$ and A/G and K/G are constant throughout X for every fixed λ on L.

In this case the zeros of F coincide with zeros of y. If $\lambda_0, \lambda_1, \lambda_2, \cdots$ are the characteristic numbers of the system (3.1), (3.3), then the characteristic numbers of the system (3.1), (3.2) are $\lambda_1, \lambda_3, \lambda_5, \cdots$. The function $y(x, \lambda_i)$ has exactly i zeros on $a < x < b$ while $F(x, \lambda_i)$ has $(i-1)/2$ zeros on this interval.

---

* *Existence theorems for implicit functions*, etc., Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 315–318.

† These Transactions, vol. 30 (1928), pp. 560–566.

Now consider the system (3.1) together with the conditions

$$\begin{aligned} &\alpha(\lambda)z(a,\lambda)-\beta(\lambda)y(a,\lambda)=0, \\ &V(b,\lambda) = 0, \end{aligned} \tag{3.4}$$

where $V(b, \lambda)=\int_a^x B(t, \lambda)\, z(t, \lambda)dt$, $B(x, \lambda)$ is summable in x on X for each fixed λ on L, continuous in λ on L for each fixed x on X, and bounded numerically for all x and λ on XL by a function $M(x)$ that is summable on X. The functions $\alpha(\lambda)$ and $\beta(\lambda)$ are continuous in λ on L and the coefficients of the system (3.1), (3.4) are assumed to satisfy conditions that are sufficient to insure the validity of existence and oscillation theorems for the system (3.1), (3.5), where

$$\alpha(\lambda)z(a,\lambda) - \beta(\lambda)y(a,\lambda) = 0, \quad z(b,\lambda) = 0. \tag{3.5}$$

The symmetry that exists between y and z together with a repetition of the arguments used in treating system (3.1), (3.2) establishes the validity of separation, existence, and oscillation theorems for system (3.1), (3.4). These theorems are analogous to Theorems V, VI, and VII. The hypotheses on G/K and A/G are replaced by the hypotheses that B/K and $-G/K$ be positive and non-decreasing functions of x on X for each fixed λ on L.

Let

$$\begin{aligned} \phi(x,\lambda) &= \alpha(x,\lambda)z(x,\lambda) - \beta(x,\lambda)y(x,\lambda), \\ \psi(x,\lambda) &= \gamma(x,\lambda)z(x,\lambda) - \delta(x,\lambda)y(x,\lambda), \end{aligned} \tag{3.6}$$

where α, β, γ, and δ are absolutely continuous functions of x on X for each fixed λ on L and continuous in λ on L for each fixed x on X. We suppose that $\alpha\delta-\beta\gamma\neq 0$ on XL and that this quantity has been made identically equal to 1 by dividing through by a non-vanishing divisor. Let $\{\alpha\beta\} =\alpha\beta'-\beta\alpha'-\alpha^2G+\beta^2K$, $\{\gamma\delta\}=\gamma\delta'-\delta\gamma'-\gamma^2G+\delta^2K$, and $\{\alpha\beta\gamma\delta\}=\beta\gamma'-\alpha\delta' -\beta\delta K+\alpha\gamma G$. We suppose that $\{\alpha\beta\}$ and $\{\gamma\delta\}$ are bounded numerically for all x and λ on XL by a summable function of x. Equations (3.1) and (3.6) yield

$$\begin{aligned} \phi' &= -\{\alpha\beta\gamma\delta\}\phi - \{\alpha\beta\}\psi, \\ \psi' &= \{\gamma\delta\}\phi + \{\alpha\beta\gamma\delta\}\psi. \end{aligned} \tag{3.7}$$

Let $\bar{\phi},\bar{\psi}$, H, and J be defined by

$$\bar{\phi} = \phi e^{\omega}, \quad \bar{\psi} = \psi e^{-\omega}, \quad H = -\{\alpha\beta\}e^{2\omega}, \quad J = \{\gamma\delta\}e^{-2\omega},$$

where

$$\omega = \int_a^x \{\alpha\beta\gamma\delta\}dt.$$

We note that H and J, as functions of x and λ, have the same properties of summability and continuity that K and G have. From (3.7) we get

$$\begin{aligned} \bar{\phi}' &= H(x,\lambda)\bar{\psi}, \\ \bar{\psi}' &= J(x,\lambda)\bar{\phi}. \end{aligned} \tag{3.8}$$

Consider the system (3.1), (3.9), where

$$\begin{aligned} \phi(a,\lambda) &= 0, \\ W(b,\lambda) &= 0, \end{aligned} \tag{3.9}$$

and $W(x, \lambda)=\int_a^x C(t, \lambda)\phi(t, \lambda)dt$, where $C(x, \lambda)$ is continuous in λ on L for each fixed x on X, summable in x on X for each fixed λ on L, and bounded numerically by a summable function of x. If we impose the same conditions on H, J, and $We^{-\omega}$ that are imposed on K, G, and A, respectively, in treating system (3.1), (3.2), we get existence and oscillation theorems for the system (3.8), (3.9). Since the zeros of ϕ coincide with those of $\bar{\phi}$, we get the same existence and oscillation theorems for system (3.1), (3.9).

As special cases of the foregoing theory we get existence and oscillation theorems for many non-self-adjoint* systems. If we let $A(x, \lambda)\equiv G(x, \lambda)$ and $\alpha\neq 0$ in system (3.1), (3.2), we get existence and oscillation theorems for the system (3.1), $\alpha(\lambda)z(a, \lambda)-\beta(\lambda)y(a, \lambda)=0$, $z(a, \lambda)=z(b, \lambda)$. Similarly, if we let $B\equiv K$ and $\beta(\lambda)\neq 0$ in system (3.1), (3.4) we get another non-self-adjoint system.

As an illustration of the foregoing work, we give the following example:

$$y' = \lambda xz,\ z' = -\lambda xy,\ y(0,\lambda) = 0,$$

$$\int_0^\pi \lambda t^2 y\, dt = 0.$$

Let λ be restricted to the interval $0<\lambda<\infty$. We note that the solution of the first three equations of this system is $y=\sin(\lambda x^2/2)$, $z=\cos(\lambda x^2/2)$. The characteristic numbers of the system are the roots of the equation $\tan(\lambda\pi^2/2)=\lambda\pi^2/2$.

* See Bôcher, loc. cit., p. 39.

4. Differential systems with boundary conditions at k points

Let $a_1, a_2, \cdots, a_k, d$ be $k+1$ distinct points of X such that $a \leqq a_1 < a_2 < \cdots < a_k \leqq b$, $a_1 < d \leqq b$. Consider the system (3.1), (4.1), where

$$\sum_{i=1}^{i=k} \alpha_i(\lambda) y(a_i, \lambda) = 0, \quad y(d, \lambda) = 0, \tag{4.1}$$

and $\alpha_1, \alpha_2, \cdots, \alpha_k$ are continuous functions of λ on L.

Theorem VIII. *If for a fixed λ on L, either $K(x) \geqq 0$ on X, or $K(x) \leqq 0$ on X and in either case the equality sign does not hold over more than a null set, then $y(x)$ can have neither a maximum nor a minimum at a point of $a < x < b$ where $z(x)$ is different from zero.*

Let c be a point of $a < x < b$ for which $z(c) \neq 0$. Since y and z are absolutely continuous functions of x on X, we can determine a positive number h such that the interval $c-h \leqq x \leqq c+h$ is a subset of X and for every x on this interval $z(x) > 0$ (or $z(x) < 0$). Now

$$y(x) = \int_a^x Kz\,dt + C = \int_{c-h}^x Kz\,dt + \overline{C}, \quad C \text{ and } \overline{C} \text{ constants},$$

is an increasing (or decreasing) function of x on $c-h \leqq x \leqq c+h$, since $Kz > 0$ (or $Kz < 0$) everywhere on this interval with the exception of a null set. It follows that $y(c)$ can be neither a maximum nor a minimum of $y(x)$.

A similar argument to the above proves that under the hypothesis that $G(x)$ be negative on X, with the possible exception of a null set of points, $z(x)$ can have extremums only at the ends of the interval X and at the points where $y(x) = 0$.

Let $y_1(x, \lambda)$, $z_1(x, \lambda)$ be the solution of system (3.1) such that $y_1(a, \lambda) = 0$, $z_1(a, \lambda) = 1$ for all values of λ on L, and let $y_2(x, \lambda)$, $z_2(x, \lambda)$ be the solution of that system that satisfies the conditions $y_2(a, \lambda) = 1$, $z_2(a, \lambda) = 0$ for all values of λ on L. Since these two solutions are linearly independent, the general solution of (3.1) can be written as a linear combination of them. Let c be any point of the interval $a < x \leqq b$; then

$$\begin{aligned} y(x,\lambda) &= y_2(x,\lambda) y_1(c,\lambda) - y_1(x,\lambda)[y_2(c,\lambda) + 1], \\ z(x,\lambda) &= z_2(x,\lambda) y_1(c,\lambda) - z_1(x,\lambda)[y_2(c,\lambda) + 1] \end{aligned} \tag{4.2}$$

is the solution of (3.1) that satisfies the condition $y(a, \lambda) + y(c, \lambda) = 0$ for every λ on L. Concerning y and z as given by (4.2) we have

THEOREM IX. *If for a fixed value of λ on L, $K(x)$ is positive, $G(x)$ is negative, K/G is a non-decreasing function of x on X and there exists a subinterval, $a \leqq x \leqq h$, of X over which K/G actually increases, then the zeros of $y(x)$ and $y_2(x)$ separate each other on X.*

The corollary to Theorem IV that is the analogue of Corollary 1 to Theorem III, together with the hypothesis that K/G actually increase throughout a neighborhood of a, yields

$$| y_2(a) | > | y_2(x_1) | \geqq | y_2(x_2) | \geqq \cdots \geqq | y_2(x_j) | ,$$

where $a, x_1, x_2, \cdots, x_j$ are the ordered zeros of $z_2(x)$ on X. Theorem VIII shows that $y_2(x)$ attains its only extremums at the zeros of $z_2(x)$, hence $|y_2(x)| < 1$ and $y_2(c)+1>0$ for every x on $a<x\leqq b$. Since y_1 and y_2 cannot both vanish at the same point, it follows from equations (4.2) and the inequality $y_2(x)+1>0$ that $y(x)$ and $y_2(x)$ cannot both vanish at the same point. Let r and s be any two zeros of $y_2(x)$ on X and assume that $y(x)$ does not vanish for any x on $\overline{X}$: $r \leqq x \leqq s$. The function y_2/y is continuous on $\overline{X}$ and has a summable derivative that is given by

$$\frac{d}{dx}(y_2/y) = KW(y_1, y_2)[y_2(c) + 1]/y^2$$

almost everywhere on X, where $W(y_1, y_2) = y_2z_1 - y_1z_2$ is the wronskian of y_1 and y_2 and is positive on X. If we write y_2/y as the indefinite integral of this derivative, we note that y_2/y is an increasing function of x on $\overline{X}$ and hence cannot vanish at both $x=r$ and $x=s$. This contradicts our hypothesis that r and s were zeros of y_2 and proves that between every two zeros of $y_2(x)$ there is at least one zero of $y(x)$. A similar argument, when applied to y/y_2, shows that between every pair of zeros of $y(x)$ there is a zero of $y_2(x)$ and thus completes the proof of our theorem.

The foregoing argument, together with an examination of (4.2), shows that when $y_1(c)>0$, $y(x)$ vanishes once between $x=a$ and the first zero of $y_2(x)$ on X and when $y_1(c)<0$, $y(x)$ does not vanish between $x=a$ and the first zero of $y_2(x)$ on X.

Let the coefficients of system (3.1) satisfy conditions that are sufficient to insure the existence of characteristic numbers, $\lambda_0, \lambda_1, \lambda_2, \cdots$ for the system (3.1), $y(a, \lambda)=1$, $y(d, \lambda)=0$, and let the hypotheses of Theorem IX be satisfied for every fixed λ on $\lambda_0 \leqq \lambda < L_2$.

THEOREM X. *There exists an infinite set of characteristic values, $k_1, k_2, \cdots$, for the system* (3.1), (4.3) *such that $\lambda_0 < k_1 < \lambda_1 < k_2 < \cdots$, where*

$$y(a,\lambda) + y(c,\lambda) = 0, \quad y(d,\lambda) = 0. \tag{4.3}$$

The solution $y(x, \lambda)$, $z(x, \lambda)$ given by (4.2) satisfies the first condition of (4.3) for all values of λ on L. Ettlinger's† implicit function theorem can be applied to show that the zeros of $y(x, \lambda)$ are continuous functions of λ on L. Let the interval X^*: $a \leqq x \leqq d$ be extended in the manner indicated under Theorem VI and let x_i be the zero of y that lies between the ith and $(i+1)$st zeros of y_2 on this extended interval. Now $x_i(\lambda_{i-1}) > d$ and $x_i(\lambda_i) < d$, since the zeros of y_2 and y move onto the interval X^* at $x = d$. For at least one value of λ between λ_{i-1} and λ_i, $x_i(\lambda) = d$. Let k_i be the set of all values of λ for which $x_i(\lambda) = d$. Since the motion of the zeros of y_2 is monotonic, it follows that $\lambda_{i-1} < k_i < \lambda_i$.

THEOREM XI. *If p_i is any value of λ in k_i, $y_2(x, p_i)$ has i zeros on $a < x < d$ while $y(x, p_i)$ has i zeros on this interval if $y_1(c, p_i) > 0$ and $i-1$ zeros on it if $y_1(c, p_i) < 0$.*

Since y_2 cannot lose any zeros at $x = a$ or at any other point of X^*, we get the first part of this theorem by a count of the zeros of y_2 as they enter X^* at $x = d$. The remainder of the theorem follows from Theorem IX.

Now consider the system (3.1), (4.1). Let all of the α's be positive and let $\alpha_1 \geqq \sum_{i=2,\ldots,k} \alpha_i$. The solution

$$y(x,\lambda) = y_2(x,\lambda) \sum_{i=1}^{i=k} \alpha_i(\lambda) y_1(a_i,\lambda) - y_1(x,\lambda) \sum_{i=1}^{i=k} \alpha_i(\lambda) y_2(a_i,\lambda),$$

$$z(x,\lambda) = z_2(x,\lambda) \sum_{i=1}^{i=k} \alpha_i(\lambda) y_1(a_i,\lambda) - z_1(x,\lambda) \sum_{i=1}^{i=k} \alpha_i(\lambda) y_2(a_i,\lambda)$$

satisfies the first condition of (4.1) for all values of λ on L. The condition $\alpha_1 \geqq \sum_{i=2,\ldots,k} \alpha_i$ insures that the coefficient of $y_1(x, \lambda)$ in $y(x, \lambda)$ is positive and this enables us to prove the same separation theorem that was proved in the preceding case.

If we replace a by a_1 in the treatment of system (3.1), (4.3) and repeat the treatment given for this system, we get the same separation, existence, and oscillation theorems for system (3.1), (4.1) that were obtained in Theorems IX, X, and XI.

It is evident that the work of this section can be applied to treat systems where z or ϕ, a linear combination of y and z, replaces y in the boundary conditions (4.1).

† Ettlinger, loc. cit.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

ANALYTIC FUNCTIONS OF HYPERCOMPLEX VARIABLES*

BY

P. W. KETCHUM

1. **Introduction.** Scheffers† has shown that many of the important theorems of monogenic functions of complex variables may be extended to functions of hypercomplex variables in algebras which are commutative and associative. Autonne‡ has studied monogenic functions of variables in non-commutative algebras. Berloty,§ basing his work on previous results of Weierstrass,‖ has also studied functions of hypercomplex variables, but his conclusions only hold for variables in Weierstrass algebras. Bechk-Widmanstetter¶ has attempted to make a practical application of Scheffer's results by trying to solve Laplace's equation in three variables.

The purpose of the present paper is to complete the work of Scheffers by extending to functions of hypercomplex variables all the important elementary theorems of complex variables, as far as is possible. Also Bechk-Widmanstetter's work is revised and extended by a study of the general problem of the solution of differential equations by analytic functions of hypercomplex variables.

2. **Preliminary definitions.** The general hypercomplex variable will be denoted by

$$w = \sum_{i=1}^{n} w_i e_i ; \quad e_i e_j = \sum_{s=1}^{n} \gamma_{ijs} e_s \qquad (i,j = 1, \cdots, n), \tag{1}$$

where the units e_i are linearly independent with respect to the complex domain, the constants γ_{ijs} are real numbers, and, except in the sections on differential equations, the w's are complex variables.

For simplicity of notation, whenever a summation is to be taken over all possible values the limits will be omitted. The index over which the sum is taken may also be omitted from under the summation sign if every index is

* Presented to the Society, September 7, 1928; received by the editors May 11, 1927.

† Leipziger Berichte, vol. 45 (1893), p. 828.

‡ Journal de Mathématiques, (6), vol. 3 (1907), p. 55. See also Encyclopédie des Sciences Mathématiques, vol. 1, part 1, p. 441; Hedrick and Ingold, these Transactions, vol. 27 (1925), p. 551.

§ Thesis, Paris, 1886.

‖ Göttinger Nachrichten, 1884, p. 395.

¶ Monatshefte für Mathematik und Physik, vol. 23 (1912), p. 257.

to be summed, or in other cases where there is no ambiguity. Also the bracket expression showing for what values of the subscripts a relation holds (as shown in equation (1)) will be omitted when every subscript not connected with a summation is to be taken for all possible values.

Multiplication is always distributive. It is commutative if and only if

$$\gamma_{ik\alpha} = \gamma_{ki\alpha}, \tag{2}$$

and associative if and only if

$$\sum_i \gamma_{ik\alpha}\gamma_{lji} = \sum_i \gamma_{li\alpha}\gamma_{jki}. \tag{3}$$

Only those algebras are considered which possess a modulus or principal unit, ϵ. A modulus exists if and only if

$$\sum \gamma_{ik\alpha}\epsilon_k = \delta_{i\alpha}\,; \qquad \sum \gamma_{ik\alpha}\epsilon_i = \delta_{k\alpha}, \tag{4}$$

where $\delta_{i\alpha}$ is Kronecker's symbol.

Division, defined as the inverse of multiplication, cannot fail for every number of the algebra. In particular it cannot fail for the modulus, the modulus being its own inverse. In a non-commutative algebra there are two kinds of division called right and left hand division. If there is some number which is not divisible by α, then α is called a nilfactor. A number α is a nilfactor if and only if

$$\Delta_\alpha = \Big|\sum_i \alpha_i\gamma_{ik\alpha}\Big| = 0, \text{ or } \Delta'_\alpha = \Big|\sum_i \alpha_i\gamma_{ki\alpha}\Big| = 0. \tag{5}$$

3. **Algebraic properties.** Geometric representation of hypercomplex numbers is the same as for complex variables, but the former are vectors in $2n$ instead of in 2 dimensions. Addition is done vectorially.

We have, as for complex numbers, the inequality for any two hypercomplex numbers

$$||\alpha| - |\beta|| \leqq |\alpha + \beta| \leqq |\alpha| + |\beta|.$$

Also,

$$\sum |\alpha_i| \leqq n^{1/2}|\alpha|.$$

Hypercomplex numbers differ from complex numbers in that the factor law is not true in general; that is, if the product of two numbers, α and β, vanishes, it does not necessarily follow as in complex variables that either α or β vanish. In fact the theorem in complex numbers that the absolute value of the product (or quotient) of two numbers equals the product (or quotient) of their absolute values, does not hold in general for hypercomplex numbers.

We can, however, establish an upper bound to the absolute value of the product as follows:

We have the relation

$$|(\alpha\beta)_s| \leqq \sum_{ij} |\gamma_{ijs}\alpha_i\beta_j| .$$

Hence

$$|(\alpha\beta)_s| \leqq \sum_{ij} |\gamma_{ijs}| \cdot |\alpha| \cdot |\beta| . \tag{6}$$

Also

$$|\alpha\beta| \leqq \sum_{ijs} |\gamma_{ijs}| \cdot |\alpha_i| \cdot |\beta_j| \leqq \sum_{ijs} |\gamma_{ijs}| \cdot |\alpha| \cdot |\beta| . \tag{7}$$

There is no simple generalization of the fact that in complex numbers the amplitude of the product (or quotient) of two numbers is the sum (or difference) of their respective amplitudes. Hence there is in general no simple rule for the calculation of powers of a hypercomplex number corresponding to DeMoivre's theorem. The ratio of the product of the absolute values to the absolute value of the product of two numbers, and the direction of the product depend only on the direction of the two numbers and not on their absolute values.

4. **Elementary functional properties.** The terms function (denoted by $f(w)$), conjugate functions, limit,* continuity, uniform continuity, are defined as in complex variables so that theorems dealing with such quantities hold for hypercomplex variables as for complex variables.†

The terms series, power series,‡ convergence, divergence, absolute convergence, uniform convergence, double series, are defined as in complex variables, and many of the elementary theorems based on these quantities hold for hypercomplex variables as for complex variables.§

A function is said to be *analytic at a point* α, if it can be expressed as the sum of a power series in $(w-\alpha)$ in the neighborhood of α. An *analytic func-*

---

* The two fundamental theorems on the sum and product of limits follow for hypercomplex variables as for complex variables. The fundamental theorem for a quotient also holds if the limit of the divisor is not a nilfactor.

† E.g. in Townsend, *Functions of a Complex Variable*, every theorem of Chap. II holds for hypercomplex variables as for complex variables.

‡ In case the algebra is not associative it may be necessary to distinguish different kinds of powers depending on the order of multiplication. In that case there would be different kinds of power series to correspond to the different kinds of powers, and the position of the constants.

§ E.g. in Townsend, loc. cit., all theorems in §§42–45 inclusive hold for hypercomplex variables. The proof of the first part of Theorem III, p. 208, is slightly changed due to the necessity of using equation (6) above. It is also to be noted that nilfactors in hypercomplex variables correspond to zero in complex variables so that β_1 cannot be a nilfactor in Theorem IV, p. 212.

tion is defined as the sum of a given power series together with all the values which can be obtained by analytic continuation of the conjugate functions of that series.

If the algebra is not commutative there will be two difference quotients defined as follows:

$$\begin{aligned} \Delta w \cdot F_r(w,\Delta w) &= f(w+\Delta w) - f(w) = \Delta f, \\ F_l(w,\Delta w)\cdot \Delta w &= f(w+\Delta w) - f(w) = \Delta f. \end{aligned} \tag{8}$$

From these we obtain two derivatives, $f'(w)$ and $'f(w)$ respectively, on taking limits as $\Delta w \to 0$.*

Differentials are defined as in complex variables so that

$$dwf' = df = 'fdw.$$

If dw is taken equal to ϵ it follows that $f' = 'f$ so that right and left hand derivatives must be equal and a distinction between them is no longer necessary.

The integral of a function, which is continuous along an arc C, along C, is defined as in complex variables by the relation

$$\begin{aligned} \int_C f(w)dw &= \lim_{\substack{m\to\infty \\ \Delta_k w \to 0}} \sum_{k=1}^{m} f(\zeta_k)\Delta_k w \\ &= \sum_s e_s \int_C \sum_{ij} f_i \gamma_{ijs} dw_j. \end{aligned} \tag{9}$$

The geometric interpretation of integration is the same as in complex variables, and most of the elementary theorems hold without change.† Also,

$$\left| \int_\alpha^\beta f(w)dw \right| \leqq \sum_{ijs} |\gamma_{ijs}| \int_\alpha^\beta |f(w)| \cdot |dw| \leqq \sum_{ijs} |\gamma_{ijs}| \, ML, \tag{10}$$

where M is the maximum value of $f(w)$, on C, and L is the length of C.

Hypercomplex numbers can be represented by points in a $2n$-dimensional space. A functional relation denotes a one-to-one correspondence between the points of one space (the w space) and the points of all or part of another space (the f space). A function will now be defined as monogenic in a $2n$-dimensional region τ if it possesses a derivative for every value of w in τ, and if the conjugate functions are analytic in τ. An inner point of such a re-

* Compare Scheffers, loc. cit. It is easily shown that the definitions are equivalent.

† E.g., in Townsend, loc. cit., Ex. 1, p. 61; equations 1, 3, 4, 6, pp. 62, 63; equation 5, p. 65, hold for hypercomplex variables.

gion τ is called a *regular point*. All points which are not regular are called *singular points*. A *monogenic function* is a function which is monogenic in some region, τ, and which is defined outside of τ by analytic continuation of the conjugate functions. As an immediate consequence of these definitions it follows that *α is a singular point of $f(w)$ if and only if $(\alpha_1, \cdots, \alpha_n)$ is a singular point of at least one of the conjugate functions.*

5. **Derivatives.** As in real variables the derivative of a constant is zero, of w is ϵ, and of a sum is the sum of the separate derivatives. Also w^2 has the derivative $2w_0$ at w_0 if and only if w_0 is commutative with all numbers of the algebra. Hence $f(w)=w^2$ possesses a derivative equal to $2w$ where and only where w is commutative with every number of the algebra.* But in order for such values of w to form a $2n$-dimensional region τ the algebra would have to be commutative. Hence *in a non-commutative algebra w^2 is not monogenic.*

In view of this fundamental difference between variables in commutative and non-commutative algebras we shall confine our attention to those that are commutative.†

The function w^3 has a derivative equal to $3w_0^2$ at w_0 if and only if $w_0(ww_0)=ww_0^2$ for every value of w. This condition would certainly be satisfied if the algebra were associative.

Scheffers (loc. cit.) has shown that *for a function to be monogenic in a region τ, it is necessary and sufficient that the conjugate functions f_i satisfy the following equations*; for all values of w in τ:

$$\frac{\partial f_s}{\partial w_k} = \sum f_i' \gamma_{iks}, \tag{11}$$

or the equations

$$\frac{\partial f_s}{\partial w_k} = \sum \epsilon_j \gamma_{iks} \frac{\partial f_i}{\partial w_j}. \tag{12}$$

He has also shown that all the derivatives of a monogenic function exist and are themselves monogenic.

6. **Associativity.** If a is any number of an associative algebra,

$$(af')dw = a(f'dw),$$

or

$$\sum a_j f_i' \, dw_k \gamma_{jlt} \gamma_{iks} e_s = \sum a_j f_i' \, dw_k \gamma_{ikt} \gamma_{jls} e_s.$$

---

* Compare Scheffers, loc. cit.

† See, however, §24 on restricted variables for further results in non-commutative algebras.

If f is monogenic this relation must hold for all values of $a_j dw_k$ so that

$$\sum_{il} f_l' \gamma_{jli}\gamma_{ika} = \sum_{il} f_l' \gamma_{lki}\gamma_{jia}.$$

By (11) this becomes

$$\sum_i \gamma_{ika} \frac{\partial f_i}{\partial w_j} = \sum_i \gamma_{jia} \frac{\partial f_i}{\partial w_k}. \tag{13}$$

Furthermore, if (13) is multiplied by ϵ_j and summed over j, equations (12) are obtained.

Hence, *in an associative algebra, in order for a function to be monogenic in a region τ it is necessary and sufficient that the conjugate functions satisfy* (13) *for all values of w in τ.**

7. **Power series expansion of monogenic functions.** Consider any function $f(w)$, which is monogenic in a region τ, whose conjugate functions are homogeneous and of the rth degree (r a positive integer) in the variables w_i. By Euler's formula,

$$\sum_k w_k \frac{\partial f_i}{\partial w_k} = r f_i. \tag{14}$$

Multiplying (11) by w_k and e_s and summing with respect to k and s gives, by virtue of (14), $f'w = rf$. But f' is also monogenic and its conjugate functions are homogeneous of the $(r-1)$st degree. Hence $f''w = (r-1)f'$. Continuing this process we get

$$r!f = [f^{(r)}w^r] \text{ where } [f^{(r)}w^r] = (\cdots((f^{(r)}w)w)\cdots w).$$

But $f^{(r+1)}w = 0$ for every value of w in τ, so that $f^{(r)}$ is a constant. Hence for values of w in τ,

$$f = [\beta w^r],$$

where β is a constant.

Now let $F(w) = \sum F_i e_i$ be any function which is monogenic in the neighborhood of a point α. Since the conjugate functions are analytic they can be expanded in power series in the variables $\{w_i - \alpha_i\}$. Thus

$$F_i = \sum_{m=0}^{\infty} f_{mi},$$

where f_{mi} are homogeneous functions of degree m in $\{w_i - \alpha_i\}$. These series

* Proved in a different manner by Scheffers (loc. cit.)

converge in the neighborhood of $(\alpha_1, \cdots, \alpha_n)$ and the derived series will also converge in the neighborhood of that point. Hence

$$\frac{\partial F_i}{\partial w_k} = \sum_{m=0}^{\infty} \frac{\partial f_{mi}}{\partial w_k}.$$

But from (12) we have in the neighborhood of $(\alpha_1, \cdots, \alpha_n)$,

$$\sum_{m=0}^{\infty} \frac{\partial f_{ms}}{\partial w_k} = \sum_{ij} \sum_{m=0}^{\infty} \gamma_{iks} \epsilon_j \frac{\partial f_{mi}}{\partial w_j}.$$

Equating terms of like degree,

$$\frac{\partial f_{ms}}{\partial w_k} = \sum \gamma_{iks} \epsilon_j \frac{\partial f_{mi}}{\partial w_j}.$$

Hence by (12), $f_m = \sum f_{mi} e_i$ are monogenic functions in the neighborhood of the point $(\alpha_1, \cdots, \alpha_n)$. But the conjugate functions f_{mi} are homogeneous and of degree m in $\{w_i - \alpha_i\}$, so by the theorem just proved $f_m = [\beta_m (w-\alpha)^m]$. Hence for values of w in the neighborhood of α,

$$F = \sum_{m=0}^{\infty} [\beta_m (w - \alpha)^m]. \tag{15}$$

This shows that any function which is monogenic in the neighborhood of a point α can be expanded in a power series in $(w-\alpha)$ which converges absolutely in some region about α. Therefore *if a function is monogenic in a given region it is also analytic in that region.**

8. **Transformation of the units.** A transformation of the units is defined by the relation

$$E_i = \sum \alpha_{ij} e_j, \qquad |\alpha_{ij}| \neq 0,$$

where α_{ij} are arbitrary complex constants. In applying such a transformation to the general variable w, we require that

$$W = \sum W_i E_i = \sum w_j e_j = w. \tag{16}$$

From this relation the variables W_i can be found as linear functions of the variables w_j. Since $W = w$, $f(W) = f(w)$ and $F(W)$ has a derivative equal to $f'(w)$ wherever $f(w)$ is monogenic. Also the conjugate functions of $f(W)$ are linear functions of the conjugate functions of $f(w)$ so that they are analytic whenever $f(w)$ is monogenic. Hence, *under linear transformation of the units, monogenic functions remain monogenic.*

* Proved by Scheffers (loc. cit., p. 846) for an associative algebra.

The following is an important transformation:

$$E_1 = \epsilon = \sum \epsilon_i e_i ; \qquad E_j \text{ arbitrary for } j \neq 1. \tag{17}$$

9. **Linear independence of derivatives of conjugate functions.** Consider the function $f(w) = [w^m]$, $m = 1, 2, \cdots$. Suppose w is transformed according to (17). Then

$$f_i = \text{const.} \cdot W_i W_1^{m-1} + \text{terms of lower degree in } W_1 ;$$

therefore

$$p_i^t \equiv \frac{\partial^t f_i}{\partial w_1^t} = \text{const.} \cdot W_i W_1^{m-t-1} + \text{terms of lower degree in } W_1 \quad (t < m-1).$$

It thus appears that *for $f(w) = [w^m]$ the derivatives p_i^t are linearly independent with respect to both i and t* ($t = 0, 1, \cdots, m-1$). The same statements can at once be made for every polynomial, $\sum_{i=0}^{m} [\alpha_i w^i]$, provided α_m is a multiple of the modulus. Also, for a fixed value of t the only terms of total degree q in p_i^t, for a function $h(w) \equiv \sum_{i=0}^{\infty} [\beta_i w^i]$, come from that term of the series which is of degree $q+t$. But for different values of i, the p_i^t are linearly independent for the function $[w^{q+t}]$. Hence *the derivatives p_i^t are linearly independent with respect to i, for every function, $h(w)$, for which there is at least one coefficient, β_i, $i > t$, which is a multiple of the modulus*; and for values of w in the region of convergence of the series.

10. **Monogenicity and analyticity.** Consider a function which is monogenic in a $2n$-dimensional region τ about a point α. By the substitution $w' = w - \alpha$ a function $f(w')$ is obtained which is monogenic in a region τ' about the origin. Since the conjugate functions of $f(w')$ remain analytic on transforming the variable according to (17), it follows that, for w in τ,

$$\frac{\partial^2 f_s}{\partial W_l \partial W_k} = \frac{\partial^2 f_s}{\partial W_k \partial W_l} . \tag{18}$$

Applying equations (12) and (17) to both members of (18) we get, for w in τ,

$$\sum_{ij} \Gamma_{iks} \Gamma_{jli} \frac{\partial^2 f_j}{\partial W_1^2} = \sum_{ij} \Gamma_{ils} \Gamma_{jki} \frac{\partial^2 f_j}{\partial W_1^2},$$

where the Γ's are the multiplication constants of the transformed units. But the derivatives p_j^2 are linearly independent for all functions, $h(w)$, which satisfy the conditions of the last section. For such functions the last equation reduces to (3) showing that the algebra must be associative. Hence *in a non-associative algebra no such function could be monogenic in τ.*

From this point on we shall confine our attention to associative algebras.

11. Fundamental theorems on nilfactors. At a regular point, α, of any analytic function $f(w)$, when dw is a nilfactor df must also be one; and if $f'(\alpha)$ is not a nilfactor, dw must be a nilfactor whenever df is such.

Also, *at a regular point α, the increment Δf becomes a nilfactor whenever the increment Δw is a nilfactor.* For there must exist a positive number δ such that, for $|\Delta w| \leqq \delta$, Δf becomes a nilfactor whenever Δw is such. If this were not so it would be possible to pick a set of points w dense at α for which the difference quotient would be infinite and hence the derivative would not exist, which is contrary to the hypothesis. But when Δf is a nilfactor the determinant $\Delta_{\Delta f}$ vanishes. Hence $\Delta_{\Delta f}=0$ whenever Δw is a nilfactor, for $|\Delta w|<\delta$. But if Δw is a nilfactor then $\rho\Delta w$ is also a nilfactor, if ρ is a complex number. But this means that Δf is a nilfactor at all points of a segment of a straight line of length 2δ, and $\Delta_{\Delta f}$ must vanish over the same segment. But $\Delta_{\Delta f}$ is an analytic function of the complex variables w_k, so that by analytic continuation, if $\Delta_{\Delta f}$ vanishes over a segment of a straight line, it vanishes over the entire line. Hence Δf is a nilfactor at all points in the plane $\rho\Delta w$.

*Every commutative and associative algebra is the direct sum of t integral subalgebras, S^i, whose units, τ_j^i, have the following properties:**

$$\tau_0^i\tau_j^i = \tau_j^i, \qquad \tau_j^i\tau_k^i = \sum\gamma_{jkl}^i\tau_l^i, \quad l>j, \quad l>k. \tag{19}$$

The variable w may be written in terms of these units as follows:

$$w = \sum_{i=1}^{t}\sum_{k=0}^{s_i} W_k^i\tau_k^i, \tag{20}$$

where s_i is the number of units in the algebra S^i. *The variables W_k^i are linear functions of the w_j's, and it is easily shown that they are linearly independent.*

Expressing $f(w)$ in terms of the units τ_j^i we get

$$f(w) = \sum_{i=1}^{t}\sum_{k=0}^{s_i} F_k^i(W_0^i, \cdots, W_{s_i}^i)\tau_k^i, \tag{21}$$

where the functions F_k^i are linear functions of the f_j's.

It is easily shown that $f(w)$ has a derivative if and only if the t functions $F^i(W^i)$ have derivatives, where

$$W^i = \sum_{k=0}^{s_i} W_k^i\tau_k^i, \quad F^i = \sum_{k=0}^{s_i} F_k^i\tau_k^i. \tag{22}$$

* Dickson, *Linear Algebras*, p. 57.

Hence *every monogenic function of w is the sum of t independent functions F^i, each of which is a monogenic function of a different hypervariable W^i.* Thus, since the component variables W^i act entirely independently of each other, their functions always remaining in the subalgebra S^i, the general problem of studying monogenic functions of hypervariables is reduced to a study of functions of hypervariables which contain only a single idempotent unit.

We shall now confine our attention (as far as §24) to monogenic functions of a hypervariable W^i. Since only one variable will be considered the superscripts will be omitted.

All values of W except multiples of the modulus, τ_0, are nilpotent and therefore are nilfactors. If s is the number of units of the variable, *the $(2s-2)$-dimensional flat, $W_0=\alpha_0$, constitutes exactly those points where $W-\alpha$ is a nilfactor. If a monogenic function vanishes at a regular point α, then the function value is a nilfactor at every point of the flat $W_0=\alpha_0$.*

Evidently *$\tau_0/(W-\alpha)$ is infinite for and only for values of W on the flat $W_0=\alpha_0$.* The same is true for $\tau_0/(W-\alpha)^k$.

12. **Linear fractional transformation.** The transformation $W'=W+\alpha$ is a translation. The transformation $W'=\beta W$ is a homogeneous strain of the entire space. If β is a nilfactor the transformation is singular, the transformed space being of lower dimensionality than the original space. If $\beta=\beta_0\tau_0$ the transformation is a mere expansion of the space, but if β has components in nilpotent directions, the space is sheared.

The points $W=W_0\tau_0$ are the only ones that remain finite under the transformation $W'=\tau_0/W$. These points behave exactly as under the corresponding transformation in complex variables.

The general linear fractional transformation is a combination of the above special cases.

13. **Extension of Cauchy's integral theorem.** *The integral of a function, $F(W)$, around a closed curve, C, vanishes, provided there exists a diaphragm* S (i.e., a region of a surface) whose complete boundary is C, such that $F(W)$ is monogenic at every point of S.*†

Generalizing Stokes' theorem to s dimensions‡ gives

---

* Only ordinary curves, surfaces, and spaces are considered in this paper.

† Just as in complex variables the curve C need not be a single curve, since cross cuts can be introduced so as to make the curve single and the diaphragm simply connected.

The proof which follows is a direct generalization of the proof in complex variables which is based on Green's theorem. See Forsyth, *Theory of Functions*, 3d edition, p. 28. Cauchy's theorem may also be proved by the method of Goursat (Whittaker and Watson, *Modern Analysis*, 3d edition, p. 85).

‡ See J. B. Shaw, these Transactions, vol. 14 (1922), p. 224.

$$(23)\qquad \int_C \sum_{i=1}^{n} P_i\{W_i\}dW_i = \iint_S \sum_{pq=1}^{n}{}^{*}\left(\pm\frac{\partial P_p}{\partial W_q}\mp\frac{\partial P_q}{\partial W_p}\right)dW_p dW_q,$$

where for the present purpose the signs are immaterial except that they must be opposite in the two terms in the bracket. The starred sum is taken over only those values of p and q for which $p<q$. The P_i and their derivatives are any continuous functions. From (9) and (23),

$$(24)\qquad \int_C F(W)dW = \sum_k \tau_k \iint_S \sum_{pq}{}^{*}\left(\pm\frac{\partial\sum_i \gamma_{ipk}F_i}{\partial W_q}\mp\frac{\partial\sum_i \gamma_{iqk}F_i}{\partial W_p}\right)dW_p dW_q.$$

But from (13) the right hand member of (24) vanishes, so the integral on the left also vanishes, and the theorem is proved.

14. **Fundamental integral theorem.** Consider the function

$$\phi(W) = \int_\alpha^W F(W)dW,$$

where $F(W)$ is any function which is monogenic in some simply connected region which includes W and α. It can be shown as in complex variables that $\phi(W)$ is single valued and monogenic and has the derivative $F(W)$.† *The fundamental theorem of the integral calculus therefore holds for hypercomplex variables*; thus,

$$(25)\qquad \int_\alpha^\beta F(W)dW = \psi(\beta) - \psi(\alpha),$$

where $\psi(W)$ is any primitive function of $F(W)$.

It is easily shown that the function W^m (m a positive or negative integer) is monogenic for all values of W, with the exception of nilfactors when m is negative, and has the derivative mW^{m-1}.‡ Hence by the fundamental integral theorem

$$(26)\qquad \int_\alpha^W W^m dW = \frac{W^{m+1} - \alpha^{m+1}}{m+1},\quad m \neq -1.$$

If $W=\alpha$ this integral vanishes, so that for any closed curve the above integral vanishes as long as $m \neq -1$.

---

† Goursat, *Mathematical Analysis*, vol. 2, part 1, p. 72. Special analysis is required in case h is a nilfactor.

‡ The sum and the product of two monogenic functions are monogenic, and likewise the quotient as long as the divisor is not a nilfactor. The formulas for the derivative of a product and of a quotient are the same as in complex variables.

15. **Converse of Cauchy's theorem: Morera's theorem.** *Consider a function $F(W)$ which is continuous in a region S of any surface in the W space. If the conjugate functions are analytic and if the integral of $F(W)$ vanishes for every closed curve lying entirely within S, then $F(W)$ is monogenic.* The proof is as in complex variables.*

16. **Singularities.** An integral around a closed curve, C, depends only on the values of the function on C and is independent of the way in which the diaphragm (as used in the proof of Cauchy's theorem) is situated. But if the contour integral does not vanish, there must be some singular points on the diaphragm, S, and hence on *every* diaphragm, S. Hence there must exist a space of $2s-2$ dimensions, which is linked† by C, and every point of which is a singular point.

If every point of the flat, $W_0=\alpha_0$, is a singular point of a function $F(W)$, and if there are no other singular points in the neighborhood of that flat, then $F(W)$ is said to have an *ordinary singularity* at α.

The *residue* of a singularity is the value of

$$\frac{\tau_0}{2\pi i}\int_C F(W)dW,$$

where C is any curve linking that singularity and no other. If the curve C lies in the plane $W=W_0$ and incloses the origin of that plane, it will link the singularity of τ_0/W, and

$$\int_C \frac{\tau_0 dW}{W} = 2\pi i\tau_0, \tag{27}$$

if C is traversed in a counterclockwise direction. Hence *the residue of the singularity of τ_0/W is τ_0.*

17. **Liouville's theorem.** *If an analytic function is bounded throughout the finite region the function is a constant.* For, if the function has no singularities the conjugate functions also have none. But each conjugate function is an analytic function of s, or less, complex variables, and such functions must be constants. Since the conjugate functions are constants the function $F(W)$ is a constant.

---

* Goursat, loc. cit., p. 78.

† The term *linked* may be regarded as defined by the above conditions, i.e., if every diaphragm, of which a closed curve C forms the complete boundary, intersects a $(2s-2)$-dimensional space, then C is said to link the space.

18. Cauchy's integral formula. *Let C be a closed curve not linking or passing through any singularities of an analytic function $F(W)$. If α is any point such that the flat $W_0=\alpha_0$ is linked by C, then*

$$F(\alpha) = \frac{\tau_0}{2\pi i}\int_C \frac{F(W)dW}{W-\alpha}. \tag{28}$$

The proof is almost exactly the same as in complex variables if the auxiliary circle γ is taken to be in the plane $W=\alpha+W_0$.

It can be shown by the classical method of real variables, if $\phi(W, \alpha)$ is an analytic function of W and α, for α in any region V, and for W in a region including any curve C, that for α in V,

$$\frac{d}{d\alpha}\int_C \phi(W,\alpha)dW = \int_C \frac{\partial\phi(W,\alpha)}{\partial\alpha}dW. \tag{29}$$

Hence

$$F^{(m)}(\alpha) = \frac{m!\tau_0}{2\pi i}\int_C \frac{F(W)}{(W-\alpha)^{m+1}}dW. \tag{30}$$

19. Infinite series. The sum of a series of continuous functions which converges uniformly in a region S is continuous in S. A series of continuous functions which converges uniformly along an arc of a curve can be integrated term by term along that arc.*

A series may also be differentiated term by term if the resulting series is a uniformly convergent series of continuous functions.

Consider a uniformly convergent series whose terms are monogenic functions. Since the series can be integrated term by term, the sum $F(W)$ is continuous and the integral of $F(W)$ vanishes for every closed curve C in the region where the terms are monogenic. But the uniform convergence of the given series requires that the series of conjugate functions converge uniformly in the corresponding region and in that case the conjugate functions of $F(W)$ are analytic. Hence from Morera's theorem *every uniformly convergent series of monogenic functions defines a monogenic function.* In particular, every power series defines a monogenic function. Hence *the necessary and sufficient condition for a function to be monogenic is that it be analytic.* A distinction between analytic and monogenic functions is no longer necessary.

---

* See Townsend, loc. cit., p. 223. Equation (7) must be employed in the proof for hypercomplex variables.

The region of convergence of a power series will now be investigated. If $W=W_0\tau_0+W_N$, where W_N is the nilpotent part of W, any power series takes the form

$$(31)\quad \sum_{m=0}^{\infty}\alpha_m W^m = \sum_{m=s}^{\infty}\alpha_m \sum_{k=0}^{s}\binom{m}{k} W_0^{m-k}W_N^k + \text{finite number of terms},$$

where s is the number of nilpotent units and $\binom{m}{k}$ is the binomial coefficient.

Suppose the series (31) converges for $W=a\tau_0$ where a is an ordinary complex constant. Then $\sum\alpha_m W_0^m$ converges absolutely and uniformly for $|W_0|<|a|$. If the conjugate series of this series are written out and differentiated successively, it follows easily that the series $\sum_{m=0}^{\infty}\alpha_m\binom{m}{k}W_0^{m-k}$ converge absolutely and uniformly for $|W_0|<|a|$. Multiplying these series by W_N^k and summing over k from 0 to s will give a series which still converges absolutely and uniformly for $|W_0|<|a|$. But the result thus obtained is the right hand member of (31), except for a finite number of terms. Hence if the series in the left member of (31) converges for $W=a\tau_0$, it converges absolutely and uniformly for all values of W for which $|W_0|<|a|$.

If the coefficients α_m are written $\alpha_m=\sum\alpha_{im}\tau_i$, then either $\alpha_m=0$ or there is a least value of i, say k, for which $\alpha_{km}\neq 0$. In this case, it is evident from the properties of the units τ_i that

$$\alpha_m W_N \equiv \sum(\alpha_m W_N)_i\tau_i = 0 \text{ for } i \geqq k.$$

Hence the numbers α_m and $\alpha_m W_N$ are either linearly independent or zero. The corresponding terms,

$$\alpha_m W_0^m\tau_0 \text{ and } \alpha_m\binom{m}{k}W_0^{m-k}W_N^k,$$

must also be linearly independent; and as a consequence, in order for (31) to diverge it is sufficient that the series $\sum\alpha_m W_0^m$ diverge. But if the latter series diverges for $W_0=a$ it also diverges wherever $|W_0|>|a|$. Hence the series (31) converges if and only if $\sum\alpha_m W_0^m\tau_0$ converges; and *there will exist a complex number a such that* (31) *diverges if* $|W_0|>|a|$ *and converges absolutely and uniformly if* $|W_0|<|a|$. A corresponding statement can be made for a power series in $W-\alpha$.

20. **Taylor's expansion.** *If $F(W)$ is analytic in a region where W is such that $|W_0-\alpha_0|<|a|$, where α is any point and a is any complex constant, then for values of W in this region,*

$$F(W) = \sum_{m=0}^{\infty} \frac{(W-\alpha)^m}{2\pi i} \int_C \frac{F(t)dt}{(t-\alpha)^{m+1}} = \sum_{m=0}^{\infty} \frac{F^{(m)}(\alpha)}{m!}(W-\alpha)^m, \tag{32}$$

where C is a closed curve within this region which links the flat $W_0=\alpha_0$, and t represents values of W on C. The proof is analogous to the standard proofs for complex variables.

If $\alpha+\beta$ is a singular point of $F(W)$ and if all points such that $|W_0-\alpha_0| < |\beta_0|$ are regular points, the power series expansion of $F(W)$ at the point α will converge where $|W_0-\alpha_0| < |\beta_0|$ and diverge where $|W_0-\alpha_0| > |\beta_0|$. *The region of convergence is thus limited by the nearest singularity in much the same way as in complex variables.*

A singularity of $F(W)$, one point of which is β, is said to be *isolated* if there exists a positive δ such that for all values of W for which $|W_0-\beta_0|<\delta$, $W_0 \neq \beta_0$, the function $F(W)$ is analytic.

Suppose that β is a point of an isolated singularity of $F(W)$. Let α be a point such that $|\beta_0-\alpha_0|<\delta/2$, $\beta_0 \neq \alpha_0$. If there exists a regular point $\gamma \neq \beta$ such that $\gamma_0=\beta_0$, then the power series expansion of $F(W)$ about the point α will converge at γ. But if the series converges at γ it must also converge at β which is contrary to the hypothesis that β is singular. Hence such a point γ cannot exist, and *every isolated singularity is an ordinary singularity.*

Since the formal laws of manipulation of series are the same as in complex variables, *the formal properties of analytic functions of a hypervariable are the same as for the corresponding function of a complex variable.* This is the principle of permanence of formal laws.

21. **Laurent's expansion.** *If a function $F(W)$ is analytic for all points of a region such that $|W_0-\alpha_0|>\delta$ and $|W_0-\alpha_0|<M$, where α is any point and M and δ are positive numbers, then within this region $F(W)$ can be expanded in a series of the form*

$$F(W) = \sum_{m=-\infty}^{\infty} a_m(W-\alpha)^m, \tag{33}$$

where

$$a_m = \frac{\tau_0}{2\pi i} \int_C (t-\alpha)^{-m-1} F(t)dt,$$

and C is any curve in the above region which links the flat $W_0=\alpha_0$.

The proof is nearly the same as that in complex variables. Laurent's

expansion affords an expression for an analytic function in the neighborhood of an isolated singularity.

22. **Poles and zeros.** If a finite positive integral value of k exists such that

$$\phi(W) = (W - \alpha)^k F(W), \tag{34}$$

where ϕ is analytic in the neighborhood of α and *not a nilfactor* at α, then $F(W)$ is said to have a *pole of order k* at α. If α is a pole of $F(W)$ then

$$\lim_{W_s \to \alpha_s} F(W) = \infty .$$

As in complex variables, if $F(W)$ is analytic at α and not identically zero, and if $F(\alpha)=0$, then $F(W)$ can be written in the form

$$F(W) = (W - \alpha)^k \phi(W), \tag{35}$$

where k is a positive integer and ϕ is analytic and not zero at α; and $F(W)$ is said to have a *zero of order k* at α. If α is a pole of order k of $F(W)$, then $\tau_0/F(W)$ has a zero of order k at α. The converse is not always true.

It may happen that the removal of a certain number, m, of zeros and poles from a function leaves it without zeros or poles, even though the function originally had more than m of them. Such a set will be called an *associated set of zeros and poles*. The poles and zeros of an analytic function are isolated, provided they form an associated set.*

For any analytic function, $F(W)$, the ith conjugate function is a function only of the variables $W_0, W_1, \cdots, W_i$. In particular the conjugate function F_0 is a function of W_0 alone. Furthermore, if there is a point α where $F(W)$ can be expressed in a series where the coefficients are ordinary complex numbers, then the conjugate function

$$F_0(W) = F(W_0). \tag{36}$$

As an example, if $\exp W$ is defined by the same series as in complex variables, the function $F_0(W) = \exp W_0$.

It is easily shown that the exponential function is an integral function and has no values which are nilfactors. The same is true for the exponential of any rational integral function of W. The method used in complex variables to prove that an integral function not having zero points can be expressed as an exponential function,† can be used in hypercomplex variables to show

* If the set is not associated the zeros and poles may not be isolated. For example, for a two-unit algebra, every one of the infinite set of points where $W_0=0$ is a zero of W^2, but an associated set of zeros would consist of the two points $A\tau_1$, and $-A\tau_1$, the two points coinciding in case A is zero.

† Goursat, loc. cit., p. 128.

that *any integral function, $F(W)$, which does not assume a nilfactor value at any point is expressible in the form*

$$F(W) = \exp G(W), \tag{37}$$

where $G(W)$ is an integral function.

Consider a rational integral function, $F(W)$, of the mth degree, in which the coefficients are ordinary complex multiples of τ_0. Then $F_0(W)=F(W_0)$; and this latter function is a polynomial in a complex variable which has m zeros $\rho_1, \rho_2, \cdots, \rho_m$. Computation shows that

$$F_1(W) = W_1F'(W_0), \tag{38}$$

so that $F_1(W)$ will vanish if $W_1=0$, regardless of the value of W_0. But if $W_1=0$,

$$F_2(W) = W_2F'(W_0)$$

which is satisfied if $W_2=0$. Continuing this process we find that $F(W)$ has the zeros $\rho_1\tau_0, \rho_2\tau_0, \cdots, \rho_m\tau_0$. Furthermore, if the ρ's are all distinct, $F'(\rho_i)\neq 0$, and $F_1(W)$ will vanish only if $W_1=0$, so that the zeros of $F(W)$ are unique. If, however, the ρ's are not all distinct, the zeros of $F(W)$ may be somewhat arbitrary. *An associated set will always consist of just m zeros, making proper allowance for their multiplicity.*

23. **Mittag-Leffler's and Weierstrass' theorems.** *Let $\alpha_1, \alpha_2, \alpha_3, \cdots$ be a sequence of points such that the $(\alpha_k)_0$'s are all distinct and*

$$|(\alpha_1)_0| \leqq |(\alpha_2)_0| \leqq |(\alpha_3)_0| \cdots, \quad \lim_{k\to\infty} |(\alpha_k)_0| = \infty.$$

There exists a function of W which is analytic everywhere except for $W=(\alpha_k)_0$, and which has an arbitrary integral function of $\tau_0/(W-\alpha_k)$ namely $G_k(\tau_0 \div (W-\alpha_k))$ as the principal part of the expansion at the point α_k. Furthermore the function is expressible in the form

$$\sum_{k=1}^{\infty}\left\{G_k\left(\frac{\tau_0}{W-\alpha_k}\right) - P_k(W)\right\}, \tag{39}$$

where the P_k's are properly chosen polynomials. The most general function with the desired singularities and principal parts is obtained by adding any integral function. The proof is nearly the same as in complex variables.

In particular, if $G_k(W)=W$, then

$$P_k(W) = -\sum_{m=1}^{r}\frac{W^{m-1}}{\alpha_k^m}, \tag{40}$$

where ν in general depends on k. If, however, there is a number p such that $\sum_k |\tau_0/\alpha_k|^p$ converges, it is sufficient to take $\nu = p-1$.

From the above theorem it follows merely by an integration that *if a function $F(W)$ has zeros of the first order at the points $\alpha_1, \alpha_2, \cdots$, it can be written in the form*

$$F(W) = e^{g(W)} \prod_{k=1}^{\infty} \left(\tau_0 - \frac{W}{\alpha_k}\right) e^{Q_\nu(W)}, \tag{41}$$

where $g(W)$ is an integral function, and ν an integer which may depend on k, and

$$Q_\nu(W) = \sum_{m=1}^{\nu} \frac{W^m}{\alpha_k^m}.$$

The proof also holds for multiple zeros, in which case the formal expression is unchanged except that some of the α's become identical.

24. **Restricted variables.** Thus far, all the variables considered have been of the type given in equation (1). The coördinates w_i may be independent functions of another set of n variables x_i without bringing about any change in the general properties of the variable. If, however, one or more relations exist between the coördinates w_i the properties are changed. Such a variable is called a *restricted variable.* If there are p such relations, p coördinates can be expressed in terms of the others so that

$$v = w = \sum_{k=1}^{n-p} w_k e_k + \sum_{k=n-p+1}^{n} h_k(w_1, \cdots, w_{n-p}) e_k. \tag{42}$$

Restricted variables are of importance in the application of hypercomplex variables in solving differential equations. The simplest example of a restricted variable is a real variable considered in the complex field.

Since a restricted variable ranges over a certain limited range of the more general variable, the properties of the restricted variable can be obtained by a consideration of the properties of the general variable within the region of restriction. *Any property of the general variable which holds for all regions alike is evidently a property of the restricted variable.* Restricted variables may, however, have properties not possessed by the general variable, as will now be shown.

We shall confine our attention to restricted variables in which the functions h_k are constants, a_k. In that case

$$v = \sum_{k=1}^{n-p} w_k e_k + \sum_{k=n-p+1}^{n} a_k e_k, \quad \text{and} \quad dv = \sum_{k=1}^{n-p} dw_k e_k.$$

It has been shown that w^2 is monogenic where and only where w is commutative with all numbers of the algebra. v^{2*} is therefore monogenic for all values of v if and only if every value of v is commutative with all numbers of the algebra. Hence *if the constants a_k are all zero, the units $e_{n-p+1}, \cdots, e_n$ need not be commutative with each other.*

For the restricted variable, v, for which the modulus is a combination of the first $n-p$ units, *the necessary and sufficient conditions that a function be monogenic are that the conjugate functions be analytic and satisfy the equations*

$$\frac{\partial f_s}{\partial w_k} = \sum_{i,j=1}^{n,n-p} \gamma_{iks}\epsilon_j \frac{\partial f_i}{\partial w_j} \qquad \begin{pmatrix} s = 1, \cdots, n \\ k = 1, \cdots, n-p \end{pmatrix}.$$

For simplicity we now restrict ourselves to commutative and associative algebras. With the same restrictions on the modulus as above, equations (13) with j and k ranging only from 1 to $n-p$, together with the analyticity of the conjugate functions, are necessary and sufficient conditions for monogenicity. Cauchy's integral theorem and his integral formula and Mittag-Leffler's theorem, as well as most of the other properties of w hold for v under restrictions that are evident in each case.

It is possible by means of restricted variables to construct a virtual exception to Weierstrass' (loc. cit.) theorem, that every algebra which allows but a finite number of zeros of polynomials is reducible to a direct sum of complex algebras. Thus, consider an algebra having the units τ_0 and τ_1. Let v be the restricted variable $v_0\tau_0$. No polynomial in v whose coefficients are not nilfactors, can have an infinity of zeros; but the algebra cannot be broken down into complex subalgebras.

25. **Applications to differential equations.** It is desired to obtain at least some real solutions of the equation

$$\Delta h(x_1, \cdots, x_p) \equiv \sum_{k=1}^{p} \alpha_k(x_1, \cdots, x_p) \frac{\partial^2 h}{\partial x_k^2} = 0, \tag{43}$$

of which Laplace's equation and the wave equation are special cases. Consider the corresponding equation

$$\Delta f(w) = 0, \quad \text{where} \quad w = \sum w_i(x_1, x_2, \cdots, x_p)e_i. \tag{44}$$

If an analytic function $f(w)$ satisfies (44), then, due to the distributive nature of Δ, the conjugate functions of $f(w)$ must satisfy (43). If, therefore, solutions of (44) are obtained, these immediately furnish real solutions of (43).†

* It should be noted that functions of a restricted variable are by no means confined to the same range of values as the variable.

† From this point on, the coördinate variables will be considered to be real.

We wish now to determine what hypercomplex variables w (of any arbitrary order n), if any, exist such that every analytic function of w satisfies (44). In such a case the conjugate functions of every analytic function of w satisfy (43). In order to determine the variable it is necessary to find values for the units (i.e., the multiplication constants, γ_{iks}) and the functions w_i. This will now be done.

From the definition of a derivative it follows that

$$\frac{\partial f}{\partial x_k} = \frac{\partial w}{\partial x_k} f'. \tag{45}$$

Hence

$$\frac{\partial^2 f}{\partial x_k^2} = \frac{\partial^2 w}{\partial x_k^2} f' + \left(\frac{\partial w}{\partial x_k}\right)^2 f'',$$

and

$$\Delta f = \sum_{k=1}^{p} \alpha_k \left[\frac{\partial^2 w}{\partial x_k^2} f' + \left(\frac{\partial w}{\partial x_k}\right)^2 f''\right] = 0. \tag{46}$$

The terms in the bracket are linearly independent for at least one analytic function f, so (46) is satisfied if and only if

$$\sum_{k=1}^{p} \alpha_k \frac{\partial^2 w}{\partial x_k^2} = 0, \text{ and } \sum_{k=1}^{p} \alpha_k \left(\frac{\partial w}{\partial x_k}\right)^2 = 0. \tag{47}$$

The first condition (47) merely requires that w be a particular solution of (44). The second condition can be written in the form

$$\sum_{ij,k=1}^{n,p} \alpha_k \frac{\partial w_i}{\partial x_k} \frac{\partial w_j}{\partial x_k} \gamma_{ijs} = 0. \tag{48}$$

From these relations and the first condition (47) (together with (2), (3), and (4)) the γ_{iks}'s and w_i's can be found. These conditions are usually not sufficient to determine uniquely a particular hypercomplex variable. In other words there is usually an infinity of variables with the desired properties. Particular variables satisfying equations (47) and (48) can be found by making certain additional assumptions. Thus if the functions w_i are assumed to be linear they automatically satisfy the first condition (47).

The particular conditions for Laplace's equation are obtained by putting $p=3$ and $\alpha_k=1$ in equations (48). An algebra whose analytic functions satisfy Laplace's equation will be called a *harmonic algebra*. If the additional assumption is made that $w_i=x_i$, equations (48) become $\sum_{k=1}^{3} \gamma_{kks}=0$. But these are the conditions that the sum of the squares of the first three units

shall vanish. Hence, under the above assumptions, *any algebra is harmonic if the sum of the squares of any three units is zero.*

For the analytic functions of a hypercomplex variable to satisfy the wave equation, when $w_1 = x_1$, the units must satisfy the relation

$$e_1^2 + e_2^2 + e_3^2 - \frac{1}{c^2} e_4^2 = 0$$

(w_4 is here taken as the time variable t).

The above solution is illustrative of the general method which can be applied to homogeneous or non-homogeneous equations of any order. The method may also be applied to simultaneous equations with one or more dependent variables. In the latter case more than one hypercomplex variable must be determined. For a linear equation with real or complex coefficients, if the w's are linear functions of the x's, the method fails in case the differential equation contains any term with derivatives of less than the second order.

26. **Solution of differential equations by the use of Scheffer's equations.** The above method of solution gives conditions which do not contain the associativity conditions. Bechk-Widmanstetter (loc. cit.) has given a method of solution of Laplace's equation, for the case where $w_1 = x_1$, which is easily generalized, and which gives equations corresponding to (48), that contain part of the associativity conditions. These equations can in fact be obtained directly from (48) by combining with (3). As will be seen, Bechk-Widmanstetter's method may also be used to determine variables for which only one or a limited number of the conjugate functions satisfy the given differential equation. This is not possible by the other method of solution.

Bechk-Widmanstetter's method will now be generalized by taking the w's to be linear functions of the x's. We assume the modulus to be e_1. This is no restriction since the modulus can always be made to satisfy this condition by a linear transformation of the units as given in (17). In any linear transformation $f(w) = f(w')$. Hence if a variable is such that its analytic functions satisfy a given differential equation, then it possesses the same property after undergoing any linear transformation. Scheffer's equations (12) now become

$$\frac{\partial f_s}{\partial w_k} = \sum_i \gamma_{iks} \frac{\partial f_i}{\partial w_1}. \tag{49}$$

But

$$\frac{\partial f_s}{\partial x_j} = \sum_k \frac{\partial f_s}{\partial w_k} \frac{\partial w_k}{\partial x_j}. \tag{50}$$

Substituting (49) into (50),

$$\frac{\partial f_s}{\partial x_j} = \sum_{ik} \gamma_{iks} \frac{\partial w_k}{\partial x_j} \frac{\partial f_i}{\partial w_1}. \tag{51}$$

Applying this relation twice, the differential equation (44) becomes equal to

$$\Delta f_s = \sum_{j=1}^{p} \sum \alpha_j \gamma_{iks} \gamma_{tqi} \frac{\partial w_k}{\partial x_j} \frac{\partial w_q}{\partial x_j} \frac{\partial^2 f_t}{\partial w_1^2}. \tag{52}$$

But in §9 it was shown that the derivatives $f_t^{(2)}$ are linearly independent. Hence, *for Δf_s to vanish it is necessary and sufficient that*

$$\sum_{j=1}^{p} \sum_{ikq} \alpha_j \gamma_{iks} \gamma_{tqi} \frac{\partial w_k}{\partial x_j} \frac{\partial w_q}{\partial x_j} = 0 \qquad (t = 1, \cdots, n). \tag{53}$$

In order for Δf to vanish it would be necessary and sufficient that equations (53) hold for $s = 1, \cdots, n$.

This method is again illustrative of a general process applicable to much more complicated differential equations. The method fails for differential equations which contain non-derivative terms.

For Laplace's equation, if we take $w_k = x_k$, equations (53) reduce to

$$\sum_{i,j=1}^{n,3} \gamma_{ijs} \gamma_{tji} = 0. \tag{54}$$

If $n = 3$ these are Bechk-Widmanstetter's conditions (loc. cit.). He has shown that they have no solution unless some of the γ's become complex. This, however, in no way impairs the usefulness of the algebra, but only indicates that n must be greater than 3.

27. **Example of a harmonic algebra.** Consider an algebra with the four units $1, i, j, ij$, which have the following multiplication table:

1	i	j	ij
i	-1	ij	$-j$
j	ij	-1	$-i$
ij	$-j$	$-i$	1

The table is easily remembered by noticing that the units i and j obey the same rules of multiplication as does the imaginary unit of complex numbers, except in respect to each other.

Now consider the restricted variable

$$w = x + aiy + bjz,$$

where i and j are units in the above table and x, y and z are the coördinates of the units. It is easily shown by the method of the sections on differential equations that every analytic function of this variable satisfies Laplace's equation provided $a^2+b^2=1$. The corresponding conjugate functions therefore represent potentials which satisfy certain particular boundary conditions. We proceed to discuss some of the simple functions of this variable. If $f(w)=U+iV+jW+ijZ$, the Scheffer's differential equations for this variable are

$$\begin{aligned}
&\frac{\partial U}{\partial x} = \frac{1}{a}\frac{\partial V}{\partial y} = \frac{1}{b}\frac{\partial W}{\partial z},\\
&\frac{\partial V}{\partial x} = -\frac{1}{a}\frac{\partial U}{\partial y} = \frac{1}{b}\frac{\partial Z}{\partial z},\\
&\frac{\partial W}{\partial x} = \frac{1}{a}\frac{\partial Z}{\partial y} = -\frac{1}{b}\frac{\partial U}{\partial z},\\
&\frac{\partial Z}{\partial x} = -\frac{1}{a}\frac{\partial W}{\partial y} = -\frac{1}{b}\frac{\partial V}{\partial z}.
\end{aligned} \tag{55}$$

28. **The function w^n.** This function may be calculated by direct multiplication. For example, for the function w^2,

$$U = x^2 - a^2y^2 - b^2z^2, \qquad W = 2bxz,$$

$$V = 2axy, \qquad Z = 2abyz.$$

The last three parts represent potentials between planes at right angles. The U part gives the potential inside and outside of an elliptic cone. If $a^2=b^2=\frac{1}{2}$ (since $a^2+b^2=1$) the cone becomes circular. To find the angle of the cone it is convenient to express U in spherical coördinates, measuring the angle θ from the X axis and ϕ from the Z axis in the YZ plane. Then

$$U = r^2[\cos^2\theta - \sin^2\theta(a^2\sin^2\phi + b^2\cos^2\phi)].$$

An electrostatic potential is constant over metallic surfaces. The function U vanishes, and is therefore constant, wherever the bracket expression vanishes, so that the metallic boundary is given by

$$\tan^2\theta = [a^2\sin^2\phi + b^2\cos^2\phi]^{-1}.$$

For $a^2=b^2=\frac{1}{2}$ this reduces to $\tan\theta=2^{1/2}$. This represents a circular cone having an angle of approximately 54°44′.

Direct calculation also shows that $\partial U/\partial\theta$ vanishes for $\theta=0$, and $\pi/2$ so that U represents the velocity potential of an elliptic stream flowing at right angles against a solid wall. If $a^2=b^2$ the stream becomes circular.

For w^3 the conjugate functions are

$$U = r^3\cos\theta[\cos^2\theta - 3\sin^2\theta(a^2\sin^2\phi + b^2\cos^2\phi)],$$
$$V = ar^3\sin\theta\sin\phi[3\cos^2\theta - \sin^2\theta(a^2\sin^2\phi + 3b^2\cos^2\phi)],$$
$$Z = 6abxyz.$$

U vanishes for $\theta=\pi/2$ and for values of θ and ϕ which make the bracket expression vanish. It represents the potential inside an elliptic cone, or the potential between an elliptic cone and a plane intersecting the cone at its vertex and at right angles to its axis. When $a^2=b^2$ the cone becomes circular and has an angle of approximately 39°14′. It also represents the velocity potential of an elliptic stream flowing into an elliptic cone along its axis and out around the sides. If $a^2=b^2$ the stream and cone are circular and the cone has an angle of approximately 63°26′.

The function V vanishes for $\phi=0$ and therefore represents the potential inside and outside of an elliptic cone which is cut in half by an infinite plane which passes through the axis of the cone. If $a^2=3b^2$ the cone is circular and has an angle of approximately 63°26′. For all functions the W part is obtained from the V part by merely interchanging ay and bz. Hence only the V part will be considered.

The Z part is the potential in a square corner.

In general for w^m the U part is the potential due to a system of concentric cones, but for $m\geqq 4$ the cones are no longer elliptical, but are of higher degree, and the constants a and b cannot be chosen so as to make them circular. The V and W parts are the potentials of a system of coaxial cones which are cut in half by a plane passing through the axis. The Z part is the same except that the cones are cut by two planes at right angles to each other, both passing through the axis of the cone.

29. **The inverse function.** A direct computation shows that

$$\Delta_w = (x^2 + a^2y^2 + b^2z^2 + 2abyz)(x^2 + a^2y^2 + b^2z^2 - 2abyz).$$

The points where $\Delta_w=0$ are nilfactors of w. In that case

$$x = 0, \text{ and } ay \pm bz = 0. \tag{56}$$

Every point of these two lines will be a point where $1/w$ becomes infinite and hence a singular point of that function. If A_{ij} is the cofactor of the term

α_{ij} in Δ_w, then the conjugate functions of $1/w$ are $U=A_{11}/\Delta_w$, $V=A_{12}/\Delta_w$ etc.

The *residue* of the second line of (56) is $\frac{1}{2}(1+ij)$, if it is linked in the direction $1, i, -1, j, 1$. The residue of the first line is $\frac{1}{2}(1-ij)$ if it is linked in the direction $1, i, -1, -j, 1$. The residue of both lines is 1 when they are linked by a curve in the x, iy plane in the direction $1, i, -1, -i, 1$ or by any curve which can be deformed into such a curve without cutting any singularities. Similarly the residue of both lines is $-ij$ when they are linked by a curve in the x, jz plane in the direction $1, j$ etc., or its equivalent.

The U part of the inverse function represents the potential due to two intersecting linear doublets. Thus, a positively charged wire lying directly above an equal negatively charged wire is situated along the first axis given in (56), and there is a similar arrangement along the other axis.

The V part differs from the U part only in that the wires lie side by side instead of one above the other, with the positive wire in both cases on the $+y$ side of the negative wire.

The Z part is the same as the U part except that the positive and negative wires are interchanged for one axis.

The function w^{-p} differs from w^{-1} only in the number of wires situated along the two axes of (56).

The exponential function is defined as in complex variables and obeys the same formal laws as in complex variables. The conjugate functions can easily be calculated directly and are found to be

$$U = e^z \cos bz \cos ay,$$
$$V = e^z \cos bz \sin ay,$$
$$Z = e^z \sin bz \sin ay.$$

These functions have two periods, one of length $2\pi/b$ in the z direction and the other of length $2\pi/a$ in the y direction. If $a=b$, the entire function has too independent periods of length $2^{3/2}\pi$. In the x direction the function behaves as in complex variables.

30. **Idempotent units.** For the more complicated functions it is convenient to employ idempotent units. For this we take $\tau_1=\frac{1}{2}(1+ij)$ and $\tau_2=\frac{1}{2}(1-ij)$. The variable then becomes

$$w = x + aiy + bjz,$$
$$= x\tau_1 + (ay - bz)i\tau_1 + x\tau_2 + (ay + bz)i\tau_2.$$

Now suppose we have given any function

$$f(\omega) = R(u,v) + iI(u,v),$$

of the ordinary complex variable $\omega=u+iv$. Then for the function $f(w)$,

$$U = R(x,ay - bz) + R(x,ay + bz),$$
$$V = I(x,ay - bz) + I(x,ay + bz),$$
$$Z = R(x,ay - bz) - R(x,ay + bz).$$

Applied to the function $\log w$ these formulas give for the conjugate functions

$$U = \tfrac{1}{2}\log [x^2 + (ay - bz)^2][x^2 + (ay + bz)^2] = \tfrac{1}{2}\log \Delta_w,$$
$$V = \arctan \frac{ay - bz}{x} + \arctan \frac{ay + bz}{x},$$
$$Z = \tfrac{1}{2}\log \frac{x^2 + (ay - bz)^2}{x^2 + (ay + bz)^2}.$$

The U part represents the potential of two negatively charged wires, one lying on the first axis of (56), and the other on the second axis. The Z part is the same except that one wire has a positive charge.

It is evident that any conjugate function of a function of w is always the sum of two corresponding two-dimensional harmonic functions placed at an angle with each other.

31. **A harmonic nilpotent variable.** A few functions of a simple nilpotent harmonic algebra, which was given by Bechk-Widmanstetter (loc. cit.), will now be discussed. Considered as a four-unit algebra it can be written in the form

1	i	k	ik
i	-1	ik	$-k$
k	ik	0	0
ik	$-k$	0	0

The variable is

$$w = x + iy + kz,$$

and

$$f(w) = U + iV + kW + ikZ.$$

For the function w^m the U and V parts are always two-dimensional functions since they never contain the variable z.

For w^3, $\quad W = 3z(x^2 - y^2), \quad Z = 6xyz;$

for w^4, $\quad W = 4z(x^3 - 3xy^2), \quad Z = 4z(3x^2y - y^3).$

In general for w^m,

$$W = mz(\text{real part of } \omega^{m-1}),$$
$$Z = mz(\text{imag. part of } \omega^{m-1}) \qquad (m = 2,3, \cdots).$$

Hence W and Z represent the potentials of a configuration of $m-1$ planes intersecting on the z axis, with an angle of $\pi/(m-1)$ between consecutive planes; and another plane coinciding with the XY coördinate plane.

Also,

$$\Delta_w = (x^2 + y^2)^2.$$

Hence the nilpotent numbers are the only nilfactors. The singularity of $1/w$ is the z axis and the residue is 1. Hence Cauchy's integral formula can be applied at all points within a cylinder whose sides are parallel to the z axis and whose directrix is the given curve γ.

32. **Completeness of solution.** In neither of the above algebras can every harmonic function be expressed as a conjugate function of an analytic function of w. In fact, the simple function $1/r$ does not occur among the conjugate functions. The conjugate functions satisfy other differential equations besides Laplace's equation. For example, in the case of the first algebra, from (55)

$$\frac{\partial^2 U}{\partial x^2} = -\frac{1}{a^2}\frac{\partial^2 U}{\partial y^2} = -\frac{1}{b^2}\frac{\partial^2 U}{\partial z^2}$$

and similar equations are readily obtained for the V, W, and Z parts.

Since the singularities of functions of hypercomplex variables never reduce to mere points in any three-dimensional flat, it seems that it would be impossible to obtain every harmonic function from any single variable.* By combining the solutions obtained from different variables, however, it is possible to obtain every harmonic function. This is the method that Whittaker has employed in obtaining a complete solution of Laplace's equation and of the wave equation.

Acknowledgement: The present paper was prepared as a doctor's thesis in physics at the University of Illinois, under the direction of Professor Jakob Kunz. The writer wishes to thank the Department of Physics for affording the opportunity to carry out this investigation, and Professor Kunz for the interest and help that he has rendered.

* Recently, however, the writer has discovered an algebra with an infinity of units which does furnish all the harmonic functions.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

CONCERNING LIMITING SETS IN ABSTRACT SPACES*

BY

R. G. LUBBEN

By a *limit point* of a point set M we mean a point which is the *limit*† of an infinite sequence of distinct points belonging to M. This notion may be generalized by regarding a *point set* as being the limit, in some sense, of a *collection* of point sets. We shall give a precise formulation of such a generalization: A point set K is said to be *the limiting set* of a collection G of point sets provided that it is the set of all points P such that there exists a sequence of points $P_1, P_2, P_3, \cdots$, and a sequence $g_1, g_2, g_3, \cdots$, of *distinct* elements of G such that P is the limit of the sequence $P_1, P_2, P_3, \cdots$, and for every n, P_n belongs to g_n.‡ Theorems concerning such limiting sets are frequently

---

* Presented to the Society, September 10, 1925, and September 6, 1928; received by the editors in June, 1926, and January, 1928. The author proved Theorems 1, 16, and 17 while holding a National Research Fellowship.

† The theorems in this paper will be stated for a space L Fréchet or for a specialization of such a space. A space L is a class of elements, which we shall call *points*, such that (1) if P is an element of L and $P_1, P_2, P_3, \cdots$ is a countable sequence of elements belonging to L, then the statement *P is the limit of the sequence $P_1, P_2, P_3, \cdots$* has a definite meaning and the question whether this statement is true or false has a determinate answer as soon as the element P and the sequence in question are themselves determined; (2) if the element P is the limit of the sequence $P_1, P_2, P_3, \cdots$, and $n_1, n_2, n_3, \cdots$ is a sequence of positive integers such that $n_1<n_2<n_3<\cdots$ then P is the limit of $P_{n(1)}, P_{n(2)}, P_{n(3)}, \cdots$ (where, for typographical reasons, $n(i)$ is used in subscripts in place of n_i); (3) if P is an element of L, P is the limit of the sequence $P, P, P, \cdots$ all of whose elements coincide with the point P. Cf. M. Fréchet, (I) *Sur quelques points du calcul fonctionnel*, Rendiconti del Circolo Matematico di Palermo, vol. 22 (1906), pp. 5–6. A space D is a space L in which with every pair of points A and B there is associated a number $d(A, B)$ such that (1) $d(A, B)=d(B, A)\geqq 0$, (2) $d(A, B)=0$ if and only if $A=B$, (3) if A, B, and C are any three elements then $d(A, B)+d(B, C)\geqq d(A, C)$, and (4) the limit of the sequence $P_1, P_2, P_3, \cdots$ is the point P if and only if $d(P_n, P)$ approaches zero as n approaches infinity. Cf. Fréchet, loc. cit., p. 30, and also (II), *Les ensembles abstraits et le calcul fonctionnel*, Rendiconti del Circolo Matematico di Palermo, vol. 30 (1910), p. 2, paragraph 3, and (III), *Sur les ensembles abstraits*, Annales Scientifiques de l'École Normale Supérieure, (3), vol. 38 (1921), p. 350, lines 1–10. The symbol $d(A, B)$ is read "the distance from A to B."

‡ Cf. R. L. Moore, *Report on continuous curves from the view-point of analysis situs*, Bulletin of the American Mathematical Society, vol. 29 (1923), p. 297. Concerning variations of this definition (mostly for less general spaces), and also of definitions, to be given later, of other types of limiting sets, see the following papers: L. Zoretti, (I), *Sur les fonctions analytiques uniformes qui possèdent un ensemble parfait discontinu de points singuliers*, Journal de Mathématiques Pures et Appliquées, (6), vol. 1 (1905), pp. 8–9, and (II), *Un théorème de la théorie des ensembles*, Bulletin de la Société Mathématique de France, vol. 37 (1909), pp. 116–119; A. Schoenflies, *Bemerkung zu meinem zweiten Beitrag zur Theorie der Punktmengen*, Mathematische Annalen, vol. 65 (1908), pp. 431–432; S. Janis-

applicable in point set theory and in abstract topology; as an example we may mention theorems concerning conditions under which the limiting set is connected (cf. Theorems 11 and 12, and the discussion and references there given).

It is of interest to discover what properties hold for limiting sets in euclidean spaces, and to determine the general classes of abstract spaces in which such theorems hold. One of the purposes of the study of abstract spaces is the discovery of conditions, preferably necessary and sufficient, that a given theorem should hold in that space. As an example we may consider the result shown in Theorem 1 of this paper. In ordinary space the limiting set of every collection of point sets is closed. We show in Theorem 1 that in a space L a necessary and sufficient condition that the limiting set of every infinite collection of point sets be closed is that the space under consideration be a space S.* Thus, the majority of the properties of a euclidean space have little bearing on the fact that in such a space the limiting set of a collection of point sets is closed.

Another example of such a study is the one we make of the distributive property. A space is said to have *the distributive property*† or *to be distributive* provided that if in that space K is a closed point set and G is a collection of point sets and each point of K belongs to some sub-set of K which is the limiting set of some sub-collection of G, then K is itself the limiting set of a sub-collection G_1 of G. If it be specified in addition that G_1 be countable, the space is said to have the *countably distributive property* or to be *countably distributive*. We show in this paper that every distributive space D is locally compact‡ and further that in order that a space D have the countably

---

zewski, *Sur les continus irréductibles entre deux points*, Journal de l'École Polytechnique, (2), vol. 16 (1912), pp. 93–94; F. Hausdorff, (I), *Grundzüge der Mengenlehre*, Leipzig, Veit, 1914, pp. 233–239 and 296–304, and (II), *Mengenlehre*, Berlin, de Gruyter, 1927, pp. 145–150; L. Vietoris, *Stetige Mengen*, Monatshefte für Mathematik und Physik, vol. 31 (1921), pp. 184–194; Zoretti-Rosenthal, *Mengen die von einem Parameter abhangen*, Encyklopaedie der Mathematischen Wissenschaften, Band II, 3, Heft 7 (II C 9a), §15, pp. 938–941.

* The *derived set* of a point set is the set of all its limit points. A *space S* is a space L in which the derived set of every point set is closed. Cf. M. Fréchet, (IV), *Relations entre les notions de limite et de distance*, these Transactions, vol. 19 (1918), p. 55. If M is a point set, M' denotes the derived set of M, and $\overline{M} = M + M'$. $M \subset N$ means that M *is a sub-set of* N.

† The term is due to R. L. Moore. The question whether a bounded euclidean space has this property was raised by W. L. Ayres during a discussion of the proof of a special case of my Theorem 13, and was settled in the affirmative by R. L. Moore.

‡ A point set is said to be *compact* provided that each of its infinite sub-sets has at least one limit point. (Cf. Fréchet I, pp. 6–7.) A space D is said to be *locally compact at a point P* or to be *compact near P* provided that P belongs to a compact open set. A point set is said to be an *open set* provided that its complement in the space under consideration is closed. A space is *locally compact* provided that it is locally compact at each of its points. A point set is *separable* provided that it

distributive property it is necessary and sufficient that it be locally compact and separable.

Certain theorems concerning limiting sets have an analogy to covering theorems. The question concerning the possibility of replacing any collection of point sets in a space by a countable sub-collection having the same limiting set is analogous to the one whether the Lindelöf property* holds in the space. This question is, however, of less interest than the following one: Under what conditions is a point set the limiting set of a given collection of point sets? (See Theorems 9, 10, 12, 13, and 14.) The distributive property is of especial interest in this connection; in a space possessing this property we have a condition which may be described roughly as follows: If the property of being the limiting set of a sub-collection of G is distributed everywhere in a closed point set K, then K itself has this property. It was for the reason just indicated that the term "distributive property" suggested itself to Moore.

Return to the consideration of a hyperspace whose elements are point sets in a given space. Condition (2) mentioned in the definition of a space L is not satisfied if we interpret "the limit of a sequence" as the limiting set of a sequence of point sets.† This objection does not apply to the sequential limiting set. The limiting set of an infinite collection of point sets is called *the sequential‡ limiting set* of this collection if and only if every infinite sub-collection of the given collection has the same limiting set; a collection of point sets is *convergent* if it has a non-vacuous sequential limiting set and is *divergent* if it has a vacuous sequential limiting set.§ In Theorems 16 and 17 we show that the aggregate of all closed point sets in an abstract space may be regarded as a space L provided that we introduce a suitable definition for "the limit of a sequence," which involves the notion of a sequential limiting set. In this connection Theorems 5, 6, and 7 are of interest; Theorem 6 shows that there exist spaces L and even spaces D in which no collection of point sets contains a convergent or a divergent sub-sequence.

Janiszewski‖ gives a definition of the limit set of a sequence of point sets;

---

contains a countable sub-set such that every point of the set either belongs to this sub-set or is a limit point of it. Cf. Fréchet III, p. 341.

* A collection of open sets is said to *cover* a point set K provided that each point of K belongs to some element of the given collection. A space is said to have the *Lindelöf property* provided that for every collection of open sets covering a point set in it there exists a countable sub-collection covering the point set. Lindelöf first showed that a euclidean space has this property; cf. his paper, *Remarques sur un théorème fondamental de la théorie des ensembles*, Acta Mathematica, vol. 29 (1905), p. 188.

† Cf. Janiszewski, loc. cit., pp. 94–95, example 1.

‡ Cf. R. L. Moore, loc. cit., and Hausdorff I, loc. cit.

§ Cf. Hausdorff I, pp. 230 and 232.

‖ Loc. cit., p. 94.

we modify his definition as follows: The *limit set* of a collection of point sets is the set of all points Q such that Q belongs to the limiting set of every infinite sub-collection of the given collection. It follows from our definition that if P is a point in the limiting set of a collection G of point sets, then P belongs to the *limit* set of some infinite sub-collection of G; we show in Theorem 6 that it is not true in general, even in a space D, that P belongs to the *sequential limiting set* of a sub-collection of G.

Janiszewski* shows that the limit set may be a proper sub-set of the limiting set, and may even be vacuous. In fact, the sequential limiting set exists if and only if the limit set is the same as the limiting set; and, in case this is true, the three are identical. The limiting set of an infinite sub-collection of a collection of point sets is a sub-set of the limiting set of the collection. It is not implied in our definitions that all the elements of a collection of point sets should be distinct; it is possible, for instance that each element should consist of a single point, and that all are the same point; in this case the limiting set of the collection would be the given point.

In concluding the introduction the author wishes to thank R. L. Moore, who indicated to one of his classes the desirability of a systematic study of limiting sets, and whose criticisms and suggestions have been of great value in the preparation of this paper.

THEOREM 1. *In order that every collection of point sets in a space L should have a closed limiting set it is necessary and sufficient that the space be a space S.*

If the condition is not sufficient, there must exist a space S in which there exists a collection G of point sets whose limiting set is not closed. Let P be a point which does not belong to K, the limiting set of G, but is the sequential limit point† of a sequence $U=Q_1+Q_2+Q_3+\cdots$ of distinct points belonging to K. Let F_n be the aggregate of all elements g of G such that, for some integer $m>n$, g contains the point Q_m. It follows that $F_{n+1}\subset F_n$. Then for some sufficiently large value N of n, F_N will contain at most a finite number of elements; otherwise, as may easily be seen, P would have to belong to the limiting set of G. Let $F=G-F_N$, $P_i=Q_{N+i}$, and $T=\sum P_i(i=1, 2, 3, \cdots)$.

It follows from the definition of P_i and the fact that P_i is a point of no element of F that there exists for each positive integer i a sequence

* Loc. cit.

† A point is the *sequential limit point* of a sequence of points M provided that it is the limit of every infinite sub-sequence of M. Cf. R. L. Moore, *On the foundations of plane analysis situs*, these Transactions, vol. 17 (1916), p. 134.

$T_i=\sum P_{ij}(j=1, 2, 3, \cdots)$ of distinct points such that (1) P_i is the sequential limit point of T_i and (2) for each integer i, P_{ij} belongs to an element g_{ij} of F such that if $m\neq n$ then $g_{mi}\neq g_{ni}$.

Let $S_1=Q_{11}=P_{11}$ and let $h_{11}=G_1$ be a definite element of F containing P_{11}. For $k>1$ let $G_k=\sum h_{ki}(i=1, 2, 3, \cdots, k)$ be a sequence of elements of F and $S_k=\sum Q_{ki}(i=1, 2, 3, \cdots, k)$ be a sequence of points selected as follows: h_{ki} is an element of $\sum g_{ji}(j=1, 2, 3, \cdots, \infty)$ with $j=n_{ki}$, where n_{ki} is the smallest positive integer x such that g_{xi} is not one of the finite number of elements of F belonging to $G_1, G_2, \cdots, G_{k-1}$; Q_{ki} is the point $P_{n(k,i),i}$.* Let $W=\sum S_i(i=1, 2, 3, \cdots, \infty)$. For every integer i, W contains infinitely many points in common with T_i, and hence T is a sub-set of the derived set of W. Since W is a sub-set of a space S, its derived set is closed. It follows that there exists a set $Q_{k(1),i(1)}, Q_{k(2),i(2)}, Q_{k(3),i(3)}, \cdots$, of distinct points of W having P as a sequential limit point. But no element of G belongs to more than one of the collections G_i, each of which contains only a finite number of elements. Hence the sequence $h_{k(1),i(1)}, h_{k(2),i(2)}, h_{k(3),i(3)}, \cdots$ contains infinitely many distinct elements of G, and P belongs to the limiting set of this sequence and thus to the limiting set of G. Thus we have demonstrated the sufficiency of the condition.

Suppose that there exists a space L in which the limiting set of every collection of point sets is closed, but which is not a space S. Then there exists in this space a point set M whose derived set has a limit point P which does not belong to M'. Let $W=P_1+P_2+P_3+\cdots$ be a sequence of distinct points belonging to M' and having P as a sequential limit point. Then for each positive integer i there exists a sequence $W_i=P_{1i}+P_{2i}+P_{3i}+\cdots$ which has P_i as a sequential limit point, is a sub-set of M, and does not contain P. Let G be a collection whose elements are P_{xy}, where x and y are independent variables, each having as range the set of positive integers. Then W is a sub-set of the limiting set of G, and so is P, since the limiting set is closed. It follows that P is a limit point of M, contrary to its definition. Thus the condition is necessary.

DEFINITION. A point P is said to be a *point of condensation*† of a point set M provided that every open set containing P contains uncountably many points of M.

THEOREM 2. *In order that it should be true for each infinite collection of point sets in a space D that at most a countable number of its elements fail to*

* For typographical reasons when n_i and n_{ij} occur as subscripts or superscripts they will be printed $n(i)$ and $n(i, j)$ respectively.

† Cf. Lindelöf, loc. cit., p. 184, and Fréchet I, p. 6.

be sub-sets of its limiting set, it is necessary and sufficient that the space be separable.

We shall show first that the condition is sufficient. Let K be the limiting set of a collection G of point sets in a separable space D. Suppose that there exists an uncountable sub-collection $\overline{G}$ of the elements of G such that no element of $\overline{G}$ is a sub-set of K. There must exist a positive number y such that there exist uncountably many elements of $\overline{G}$ containing points whose distances* from K are greater than y. Let E be the set of all points belonging to elements of $\overline{G}$ and having distances from K greater than y. If a point of E is common to infinitely many elements of $\overline{G}$ it belongs to the limiting set of G. This is clearly impossible. Further, if E contains at most a countable number of points, it follows from what we have just proved that at most a countable number of elements of $\overline{G}$ contain points in E; this involves a contradiction with our definition of E. Hence E contains uncountably many points, and has a point of condensation.† Since no point of E belongs to more than a finite number of elements of $\overline{G}$, this point of condensation belongs to the limiting set of G. Since its distance from K is not less than y we have a contradiction. Hence $\overline{G}$ does not exist, and the condition given in the statement of the theorem is sufficient.

We shall show next that the condition is necessary. Suppose that it is not, and that there exists a non-separable space D in which every infinite collection of point sets has the property mentioned in the statement of the theorem. Then the space contains an uncountable point set H having no limit point.‡ Let G be a collection of elements g such that each g consists of a single point belonging to H and each point of H belongs to a single element of G. It follows that H' is the limiting set of G. From the fact that all except a countable number of the elements of G are sub-sets of H' it follows that H' is non-vacuous. Thus, the supposition that the space is not separable has led to a contradiction.

DEFINITION. A space is said to have the *Cantor-Bendixson property* provided that every closed point set in it is the sum of a perfect set and a countable set. The fact that this property holds in a euclidean space is, as

---

* If P is a point and N is a point set, by $d(P, N) = d(N, P)$, *the distance from P to N*, is meant the lower bound of the set of values $[d(P, Q)]$, where Q is a point of N.

† Cf. W. Gross, *Zur Theorie der Mengen in denen ein Distanzbegriff definiert ist*, Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse, vol. 123 (1914), IIa, pp. 805–806, and Fréchet I, p. 6.

‡ Cf. Gross, loc. cit., pp. 805–806.

indicated by Lindelöf, loc. cit., the most important part of a theorem by Cantor* and Bendixson.

THEOREM 3. *In order that a space D should have the Cantor-Bendixson property it is necessary and sufficient that the space be separable.*

The sufficiency of the condition follows from the work of Fréchet† and Gross†; the argument given by Fréchet is similar to one given for the euclidean case by Lindelöf (loc. cit., pp. 183, 187). The necessity of the condition follows easily from Gross's work. Suppose that there exists a space D which has the Cantor-Bendixson property but is not separable. Then it contains an uncountable point set M which has no limit point;‡ then M is closed. It follows that M is the sum of a perfect set and a countable set. Since M contains uncountably many points, this perfect set is non-vacuous. Every point of the perfect set is a limit point of this set and of M. Thus we get a contradiction.

THEOREM 4. *In order that every infinite collection of point sets in a space D should contain a countable sub-collection having the same limiting set as the collection itself it is necessary and sufficient that the space be separable.*

It was pointed out by W. L. Ayres that the sufficiency of the condition for a bounded euclidean space is established as a consequence of the fact that such a space has the countably distributive property. The same conclusion holds for any space D that has this property. Not all separable spaces D have this property, however; cf. Theorem 9.

Let G be a collection of point sets in a separable space D, and K be its limiting set. Since the space is separable there exists a countable§ sequence $P_1+P_2+P_3+\cdots$ every element of which is a point of K, and such that every point of K is the sequential limit point of a subsequence of it. For each positive integer n there exists an element g_n of G whose distance from P_n is less than $1/n$ and which is distinct from $g_1, g_2, g_3, \cdots, g_{n-1}$, the elements associated respectively with $P_1, P_2, P_3, \cdots, P_{n-1}$. Then K is the limiting set of $g_1+g_2+g_3+\cdots$.

The necessity of the condition may be established as follows: Let G be a collection of point sets in a space D such that (1) each element of G is a

---

* Cf. G. Cantor, *Fondements d'une théorie générale des ensembles*, Acta Mathematica, vol. 2 (1883), pp. 405, 409–414; and I. Bendixson, *Quelques théorèmes de la théorie des ensembles de points*, Acta Mathematica, vol. 2 (1883), pp. 415–427.

† Cf. Fréchet I, p. 19, and Gross, loc. cit., pp. 805–806, 815–816.

‡ Cf. Gross, loc. cit., pp. 805–806.

§ Cf. Gross, loc. cit., pp. 805–806, and Fréchet I, p. 27.

single point of the space and (2) for each point x of the space there exists a countable sequence $g_{1x}, g_{2x}, g_{3x}, \cdots$ of elements of G, each of which is the point x. Clearly the limiting set of G is M, the set of all points in the space. By our hypothesis G contains a countable sub-collection G_1 which has the same limiting set. It follows from our definition of a limiting set that if y is a point of M and is not the limit of an infinite sequence of distinct points belonging to M_1, the set of all points belonging to elements of G_1, it is the limit of a sequence $y, y, y, \cdots$ of points of M_1, each of which is the point y. Then y belongs to M_1. Since M_1 is countable, M is separable.

THEOREM 5. *If, in a space D, the limiting set of a collection G of point sets is separable, then every infinite sub-collection of G contains a sub-collection that has a sequential limiting set.*

The question as to the truth of this theorem for a bounded euclidean space was raised by R. L. Moore, and in June, 1925, was proved independently by him and by me; he pointed out that my method of proof has an advantage over his in that it applies to abstract spaces. After submitting this paper for publication I learned that T. Ważewski has the following theorem in his thesis:* In a compact space D every infinite collection of point sets contains a convergent sub-collection. Also, since this paper was submitted for publication Zarankiewicz† and Hausdorff have published analogous results, Zarankiewicz showing that in a euclidean space and Hausdorff that in a separable space D every infinite collection of point sets contains a sub-collection which has a sequential limiting set. The argument given by Hausdorff does not suffice to prove the theorem we have stated, since he uses the fact that the complement of the limiting set is separable; however, if the proof given by Zarankiewicz, which is similar to the one discovered by the author, is modified slightly, it suffices for our purposes.

In 1909 Zoretti‡ proved the following theorem: If, in a plane, G is a collection of continua§ (1) whose sum is bounded, (2) whose limit set is non-vacuous, and (3) whose limiting set is a continuum, then G contains an in-

---

* *Sur les courbes de Jordan ne renfermant aucune courbe simple fermée de Jordan*, Krakow, 1923, p. 24.

† Cf. Zarankiewicz, *Sur les points de division dans les ensembles connexes*, Fundamenta Mathematicae, vol. 9 (1927), pp. 127–129; and Hausdorff II, pp. 147–148.

‡ Cf. Zoretti II.

§ A *continuum* is a closed and connected point set. (Cf. Cantor, loc. cit., p. 406.) A point set is *connected* provided that it is not the sum of two mutually exclusive, non-vacuous point sets, neither of which contains a limit point of the other. Cf. N. J. Lennes, *Curves in non-metrical analysis situs with an application in the calculus of variations*, American Journal of Mathematics, vol. 33 (1911), p. 303, and Hausdorff I, p. 244.

finite sub-collection which has a closed and connected sequential limiting set. The conditions (2) and (3) are unnecessary, for the existence of a sequence having property (2) follows from (1), and (3) follows from (2) by a theorem of Zoretti's, p. 116, loc. cit. In his process of selecting a sequence which has a sequential limiting set, Zoretti makes no use of the fact* that the elements of his collection are continua, and this process is adequate to prove that every infinite collection of point sets in a plane whose sum is bounded contains a convergent sub-collection.

Proof of Theorem 5. Let K be a separable point set in a space D and G be a collection of point sets having K as a limiting set. Then K has a countable sub-set H such that if X is a point of K it is either a point of H or a limit point of H. For each point of H consider all the spheres† having this point as a center and having radii whose lengths are rational numbers. Let $R_1, R_2, R_3, \cdots$ be the set of the interiors of these spheres. Now proceed as in the paper of Zarankiewicz (loc. cit., pp. 128–129).

THEOREM 6. *In a space L which contains a point set having the power of the continuum and having a vacuous derived set there exists an infinite collection of point sets which has no convergent or divergent sub-collection.*

The following example shows that there exist spaces D which satisfy the hypothesis of the theorem: Let D_1 be an aggregate of objects having the power of the continuum in which all the requirements of a space D are satisfied, subject to the condition that the distance between two distinct elements of D_1 is unity.

Proof of Theorem 6. Consider a space L in which there exists a point set M having the power of the continuum, but having no limit point. Let X be the set of all real numbers x such that $0<x<1$. Then there exists a correspondence such that to each x there corresponds exactly one point P_x of M and conversely. Let the number x be expanded as an *infinite* decimal in the binary scale. Let G be a collection of point sets, $g_1, g_2, g_3, \cdots$, where the set g_n consists of all points of M for which the corresponding elements in X, when given the binary expansion just mentioned, have as their nth digit unity.

Let $T=g_{n(1)}+g_{n(2)}+g_{n(3)}+\cdots$ be a sub-collection of G such that $n_1<n_2<n_3<\cdots$. Let $z=\sum 2^{-n(2i)}$ $(i=1, 2, 3, \cdots)$, and $H=\sum g_{n(2i)}$ $(i=1, 2, 3, \cdots)$. Then P_z is common to all the elements of H and hence belongs to the limiting set of H. Hence the limiting set of T is non-vacuous, and T

* This fact is used in proving that the limiting set is connected.

† The terms *sphere, interior* and *exterior of a sphere*, etc. are defined precisely as in euclidean geometry. Cf. Fréchet I, p. 21, paragraph 34. A *spherical region* is the interior of a sphere.

is not divergent. It follows from our definitions of M and G that a point of M belongs to the limiting set of an infinite sub-collection of G if and only if the point belongs to infinitely many elements of this sub-collection. Hence P_s does not belong to the limiting set of $T-H$. Hence T does not converge.

THEOREM 7. *If the hypothesis of the continuum* is correct, then in order that every infinite collection of point sets in a space D should contain a convergent or a divergent sub-collection, it is necessary and sufficient that the space be separable.*

The theorem is a consequence of Gross, pp. 805-806, Fréchet I, p. 27, and Theorem 6, and Hausdorff II, pp. 147–148.

THEOREM 8. *Every space D which has the distributive property is locally compact.*

If the theorem is not true, there exists a space D having the distributive property, and containing a point P near which it is not compact and a countably infinite set of points, K, which has no limit point and does not contain P. Let R_1 be a spherical region containing P such that $\overline{R}_1$ contains no point of K. There exists in R_1 a countable point set M_1 having no limit point and not containing P; within R_1 there is a spherical region R_2 with its center at P, of diameter less than one half, and containing no point of M_1. This process may be continued; in general, let R_n be a spherical region of diameter less than $1/n$, which contains no point of the point sets $M_1, M_2, M_3, \cdots, M_{n-1}$, has its center at P, and lies within the region R_{n-1}; let $M_n=P_{1n}+P_{2n}+P_{3n}+\cdots$ be a countable sequence of points which lies within R_n, does not contain P, and has no limit point. Let $K=P_1+P_2+P_3+\cdots$. For every pair of integers i and j let $g_{ij}=P_{ij}+P_j$. Then P_j is the limiting set of $\sum g_{ij}$ $(i=1, 2, 3, \cdots, \infty)$. From the fact that the space under consideration has the distributive property and that K is closed, since it has no limit point, it follows that K is the limiting set of a sub-collection N of $\sum\sum g_{ij}$ (i and $j=1, 2, 3, 4, \cdots, \infty$). Then there exists an integer m such that R_m contains no points belonging to elements of N. It follows that N is a sub-collection of $\sum\sum g_{ij}$ $(j=1, 2, 3, \cdots, m-1;\ i=1, 2, 3, \cdots, \infty)$. There exists a spherical region R which contains P_{m+1} but contains no point of $K-P_{m+1}$ or of $\overline{R}_1$. But P_{m+1} does not belong to any point sets in the collection $\sum\sum g_{ij}$ $(j=1, 2, 3, \cdots, m-1;\ i=1, 2, 3, \cdots, \infty)$. It follows

* The *hypothesis of the continuum*, as formulated by Cantor, is the following: No uncountable point set has a power less than that of the continuum. Sierpiński, *Sur l'hypothèse du continu*, Fundamenta Mathematicae, vol. 5 (1924), p. 177–187, gives a discussion of a number of interesting theorems whose truths follow from that of this hypothesis.

that P_{m+1} does not belong to the limiting set of N. Thus, the supposition that the space is not locally compact has led to a contradiction.

THEOREM 9. *In order that a space D should have the countably distributive property it is necessary and sufficient that it be locally compact and separable.*

First we shall show the sufficiency of the condition. Let K be a closed point set in a separable, locally compact space D and G be a collection of point sets such that each point of K belongs to a sub-set of K which is the limiting set of some infinite sub-collection of G. Let M be the limiting set of G and $N=M-K$. The point set K is separable* and hence there exists a countable sub-set $H=P_1+P_2+P_3+\cdots$ of K such that $K \subset \overline{H}$. For each integer i there exists a sub-collection $G_i=g_{1i}+g_{2i}+g_{3i}+\cdots$ of G such that the *limit set* of G_i is a sub-set of K and contains P_i.

From the fact that K is closed and that the space is locally compact it follows that for each point Q of N there exists a spherical region R_Q which is compact and such that $\overline{R}_Q$ contains no point of K. From the fact that the space is separable and hence has the Lindelöf property† it follows that there exists a countable sequence $R_1, R_2, R_3, \cdots$ of these regions covering N. If i and j are integers, the region R_j contains points of at most a finite number of elements of G_i; for otherwise, since $\overline{R}_j$ is compact, it would contain a point of the limiting set of G_i, thus contradicting the fact that the limiting set of G_i is a sub-set of K and $\overline{R}_j$ contains no point of K. Hence, for every integer i the sequence G_i contains an infinite sub-sequence F_i such that no element of F_i contains a point in any of the regions $R_1, R_2, R_3, \cdots, R_i$. Let $F=\sum F_i$ $(i=1, 2, 3, \cdots, \infty)$. We shall show that E, the limiting set of F, is a sub-set of K. If it is not, it must contain a point W of N, since F is a sub-collection of G and E is then a sub-set of M. There exists an integer m such that W belongs to R_m. Let $n \geqq m$. Then it follows from the definition of F_n that F_n contains no point within R_m. Further, as we have pointed out above, if $1 \leqq p \leqq m$, at most a finite number of the elements of G_p, and hence of F_p, contain points within R_m. Hence R_m contains points of at most a finite number of elements of F, and hence W cannot belong to the limiting set of F.

We shall show next that K is a sub-set of the limiting set of F. It follows from the definition of F that H is a sub-set of the limiting set of F. If X is a point of $K-H$, it is a limit point of H. By Theorem 1, X is then a point of the limiting set of F. From $E \subset K$ and $K \supset E$ it follows that $K=E$.

Given a space D which has the countably distributive property. If G

* Cf. Fréchet I, p. 27, and Gross, pp. 805–806.

† Cf. Gross, pp. 805, 806, 809.

is a collection of point sets, it follows by Theorem 1 that its limiting set is closed. By our hypothesis G can be replaced by a countable sub-collection having the same limiting set. By Theorem 4 the space is separable. By Theorem 8 the space is locally compact. Hence, the condition is necessary.

DEFINITIONS. If M and N are two point sets in a space D, by $l(M, N)$, *the lower distance from M to N*, is meant the lower bound of the values $d(P, N)$, where P is a point of M; and by $u(M, N)$, *the upper distance from M to N*, is meant the upper bound of the values $d(P, N)$. If G and H are collections of point sets, we shall say that H is *upper semi-continuous with respect to G* provided that if g is a fixed element of G and h is a variable element of H, then as $l(h, g)$ approaches zero so does $u(h, g)$. If the collection G is upper semi-continuous with respect to itself, it is said to be an *upper semi-continuous* collection.*

THEOREM 10. *If in a space D, K is a closed point set and G is a collection of point sets, then a sufficient condition that K be the limiting set of G is that K is a sub-set of the limiting set of G and* (1) *contains points in common with the limiting set of every infinite sub-collection of G that has a non-vacuous limiting set*; (2) *G is upper semi-continuous with respect to the point set K. If it be specified in addition that the space be compact, the condition mentioned above is both necessary and sufficient.*

It is easy to construct examples showing that the condition mentioned in the theorem does not remain necessary in case the stipulation that the space be compact be removed. The theorem remains true if we replace (1) by (1′) following: (1′) *If e is a positive number, there exist at most a finite number of elements g of G such that $l(g, K) > e$.*

Proof of Theorem 10. Let H be the limiting set of G. If H contains a point x not belonging to K, there must exist an infinite sub-sequence $G_x = g_1 + g_2 + g_3 + \cdots$ of G such that every element of this sequence is at a distance from x less than $\frac{1}{2}d(x, K) > 0$, and for each positive integer i, $d(x, g_i) < 1/i$. By the hypothesis there exist (1) a positive number d such that when $l(g, K) < d$ then $u(g, K) < \frac{1}{2}d(x, K)$, and (2) an infinite sub-sequence of elements of G_x, every element of this sub-sequence being at a lower distance from K less than d. Thus the supposition that the point x exists has led to a contradiction, and $K = H$.

* For definitions of lower distance, of upper distance, and of upper semi-continuous collections of point sets, cf. R. L. Moore, *Concerning upper semi-continuous collections of continua which do not separate a given continuum*, Proceedings of the National Academy of Sciences, vol. 10 (1924), p. 356, and *Concerning upper semi-continuous collections of continua*, these Transactions, vol. 27 (1925), p. 416.

We shall now show that if the space is compact the condition is necessary. Part (1) of the condition follows immediately. If part (2) does not follow, there must exist a positive number ϵ and a sub-sequence $H=H_1+H_2+H_3+\cdots$ of elements of G such that, for every n, $u(H_n, K)>\epsilon$, but $l(H_n, K)$ approaches zero as n approaches infinity. Since the space is compact it follows that the limiting set of H contains a point x whose distance from K is not less than ϵ. But the result just stated involves a contradiction of our assumption that K is the limiting set of G and hence contains the limiting set of H.

DEFINITIONS. If x and y are a pair of points, a sequence of points $x_1(=x)$, $x_2, x_3, \cdots, x_n(=y)$ is said to be *a chain joining x and y*; a pair of consecutive points, x_j and x_{j+1}, is said to be a *link* of this chain and $d(x_j, x_{j+1})$ is the *length* of this link. If g is a point set let $f(g)$ be the lower bound of the set of all positive numbers h such that any pair of points in g can be joined by a chain every point of which belongs to g and every link of which has a length less than or equal to h. If g is a bounded point set, $f(g)$ certainly exists; otherwise it need not. If G is *zusammenhängend** in the sense of Cantor, $f(g)=0$.

Let M be a point set in a space L and P and Q be points of M. Let $C_{PQ}=\sum\sum P_{ij}$ $(j=1, 2, 3, \cdots, \infty; i=1, 2, 3, \cdots, n_j)$ be a sub-set of M such that for each positive integral value of j, $P_{1j}=P$, n_j is a positive integer, and $P_{n(j),j}=Q$. The points P and Q are said *to be well chained in M* provided that there exists at least one C_{PQ} such that if $j_1<j_2<j_3<\cdots$ and the sequence $\sum P_{m(j(k)),j(k)}$ $(k=1, 2, 3, \cdots, \infty)$, of the set C_{PQ} has a sequential limit point T, then T is a sequential limit point also of every infinite sub-sequence of $\sum P_{m(j(k))+1,j(k)}$ $(k=1, 2, 3, \cdots, \infty)$ and of $\sum P_{m(j(k))-1,j(k)}$ $(k=1, 2, 3, \cdots, \infty)$, where $1\leqq m_{j(k)}\leqq n_{j(k)}$. If every pair P and Q of points of M is well chained in M, M is said to be *well chained*.

THEOREM 11. *If, in a locally compact space D, $G=[g]$ is a collection of point sets, K is the limiting set of G, and $f(g)$ approaches zero as $l(g, K)$ approaches zero, then in order that K be well chained it is necessary and sufficient that there exist a well chained sub-set K_1 of K which contains points in common with every sub-set of K which is the limiting set of a sub-collection of G.*

Janiszewski considers an analogous question (1) for a bounded number space and (2) in an abstract space which may be characterized as a perfectly compact, connected space in which a distance is defined (cf. Janiszewski, p. 162, loc. cit.); he shows that in such spaces a sufficient condition that

* Cf. Cantor, loc. cit., p. 406, and *Ueber unendliche, lineare Punktmannigfaltigkeiten*, Mathematische Annalen, vol. 21 (1883), p. 576.

the limiting set be *connected* is that the *limit set exist.* Our theorem is more general than his in that we give necessary and sufficient conditions and that our space is more general.

Schoenflies (loc. cit.) gives a less general case than Janiszewski; he assumes that the sequential limiting set exists. See also Zoretti I, pp. 8–9, and Zoretti II. In connection with Zoretti's work see our comments following Theorem 5. See also Vietoris, loc. cit., p. 186, and Hausdorff I, pp. 301–302.

It might be pointed out that in a non-compact space D the limiting set of a collection of connected point sets need not be connected, cf. Hausdorff I, p. 239, last four lines. As an example of a bounded space D where the condition given in Theorem 11 is not sufficient consider the space D_3 given below. Note, however, that this space is not locally compact. Let H_x be the straight line interval in the euclidean plane having as end points $(x, 0)$ and $(x, 1)$. Let A be the point $(0, 0)$ and B the point $(0, 1)$. Let the points of D_3 be the points in the euclidean sense of $A+B+\sum H_{1/i}$ $(i=1, 2, 3, \cdots, \infty)$, and distance and limit be defined as in ordinary space. Let G be a collection of point sets whose elements are the H's in $\sum H_{1/i}$ $(i=1, 2, 3, \cdots, \infty)$. Then G is a collection of connected point sets having as a sequential limiting set the sum of A and B. The set $A+B$ is neither connected nor well chained.

Proof of Theorem 11. Obviously the condition is necessary; we shall prove its sufficiency.

First we shall show that if M is a connected point set it is "zusammenhängend." For, if it is not, there exists a positive number d and a pair of points x and y of M which cannot be joined by a finite chain of points belonging to M, each link in the chain having a length less than d. Let N be the set of all points of M which can be joined to x by a chain such as that just mentioned. Then both N and $M-N$ are non-vacuous and, since M is connected, there exists a point z which belongs to one of the point sets N and $M-N$ and is a limit point of the other. Let E be a spherical region with center at z and having a radius of length $d/3$. It follows that E contains points of both N and $M-N$. Clearly this involves a contradiction. Evidently M is well chained.

We shall show next that if M and N are two non-compact, connected point sets their sum is well chained. Let $m_1, m_2, m_3, \cdots$ and $n_1, n_2, n_3, \cdots$ be two infinite sequences of points belonging to M and N respectively, and neither of which has a limit point. It follows from the preceding paragraph that for each positive integer i there exists a pair of chains of points $m_{1i}(=m_1)$, $m_{2i}, m_{3i}, \cdots, m_{p(i),i}(=m_i)$ and $n_{1i}(=n_1), n_{2i}, n_{3i}, \cdots, n_{q(i),i}$ $(=n_i)$ be-

longing to M and to N respectively such that each link in each of these chains has a length less than $1/i$. Let C_i be the chain $m_{1i}, m_{2i}, m_{3i}, \cdots, m_{p(i),i}, n_{q(i),i}, \cdots, n_{2i}, n_{1i}$. It follows from the definitions of $C_1, C_2, C_3, \cdots$ that m_1 and n_1 are well chained in M, and since they may be selected at pleasure it follows that $M+N$ is well chained.

Hence, to prove that K is well chained it suffices to prove that either (1) K is connected or (2) every maximal* connected sub-set of K is non-compact. Suppose that neither of these conditions is realized. Then K has a compact maximal connected sub-set T which does not contain all of K. About each point of T there exists a spherical region such that the sum of its points and limit points is compact. The point set T can be covered by a finite† sub-collection of this collection of regions. Let H be the sum of the points belonging to this finite set of regions and B be the boundary‡ of H. Evidently B contains no point of T. Since $\overline{H}$ is the sum of a finite number of compact point sets, it is compact. Let G_2 be a sub-collection of G having a *limit* set containing a point P of T. Suppose that K_2, the *limiting* set of G_2, contains points without $\overline{H}$; let Q be such a point. Let d be a positive number which is less than each of the numbers $d(P, h)$ and $d(Q, H)$.§ It follows from the hypothesis of the theorem and the definition of G_2, that G_2 contains a subsequence $g_1, g_2, g_3, \cdots$ such that for each positive integer n, $f(g_n)$, $d(P, g_n)$, and $d(Q, g_n)$ are each smaller than the smaller of $d/2$ and $1/n$. For each n there exists a finite chain $F_n = x_{1n}, x_{2n}, x_{3n}, \cdots, x_{i(n)n}$ of points belonging to g_n such that (1) all except the last point in this chain belong to $\overline{H}$ but the last does not, (2) the distance of x_{1n} from P is less than $1/n$, and (3) each link has a length less than $1/n$. The limiting set F of $\sum F_n$ $(n=1, 2, 3, \cdots, \infty)$ is closed and connected,‖ contains P, and is a sub-set of T, which is, by definition, a maximal connected sub-set of K. The point set $\sum x_{i(n)-1,n}$ $(n=1, 2, 3, \cdots, \infty)$ is compact and has the same limit points as $\sum x_{i(n),n}$ $(n=1, 2, 3, \cdots, \infty)$. These points are common to K_2, T, and B. But B and T have no points in common, and hence $K_2 \subset \overline{H}$.

Further $K_2 \subset T$. For, if e is greater than zero, we can evidently cover T with a set of spherical regions each of which is compact and has a diameter less than e. In terms of these regions we can then define a point set $\overline{H}_e$ which has properties similar to those of $\overline{H}$, and which contains no point whose

---

* *A maximal sub-set, having a property P*, of a point set M, is a sub-set N of M which has the property P but is not a proper sub-set of any sub-set of M which has this property.

† Cf. Gross, p. 810, and Fréchet I, pp. 22 and 26.

‡ That is, B is the set of all limit points of H that do not belong to H.

§ Let h be the complement of $\overline{H}$.

‖ Cf. Janiszewski, loc. cit., p. 98, lines 1–12.

distance from T is greater than e. As above, $K_2 \subset \overline{H}_e$. Then $K_2 \subset T = \prod \overline{H}_{1/i}$ $(i=1, 2, 3, \cdots, \infty)$.

It follows from the hypothesis of the theorem that K_2, and hence T, contains points of K_1. By an argument analogous to that used in showing that $K_2 \subset T$ we can show that $K_1 \subset T$.

If there exists a point Y of $K-T$, there exists an infinite sub-collection G_Y of G having a limit set containing Y. It follows from the hypothesis of the theorem that G_Y contains an infinite sub-collection having a limit set containing a point of K_1. We have shown above that the limiting set of such a collection is a sub-set of T. Hence the point Y does not exist.

From $T \subset K$ and $K \subset T$ it follows that $K=T$. But this contradicts our supposition that K is not well chained. Thus we have established the truth of our theorem.

THEOREM 12. *In order that a compact point set in a locally compact space D should be the closed and connected limiting set of a collection of well chained point sets, it is necessary and sufficient that it* (1) *be a maximal well chained sub-set of the limiting set of this collection and* (2) *contain points in common with the limiting set of every infinite sub-collection of the given collection which has a non-vacuous limiting set.*

It is a consequence of the definition of well-chainedness that if P is a limit point of a point set M then $M+P$ is well chained. Let K be the limiting set of a collection of point sets and K_1 be a maximal well chained sub-set of K. By Theorem 1, K is closed. Since K_1 is a *maximal* well chained sub-set of K, it is closed. The argument used in proving Theorem 11 shows that the condition is sufficient. Its necessity is obvious.

If we leave off the condition that the space D under consideration be locally compact, the condition given in Theorem 12 is not sufficient. Consider, for example, the space D_3 mentioned in the discussion following the statement of Theorem 11. The point A satisfies the condition but is not the limiting set of any sub-collection of $\sum H_{1/i}$ $(i=1, 2, 3, \cdots, \infty)$. The point just raised applies also to each of the following two theorems.

THEOREM 13. *If $G=[g]$ is a collection of point sets in a separable, locally compact space D, K is the limiting set of G, $f(g) \to 0$ as $l(g, K) \to 0$, and K_1 is a maximal well chained sub-set of K, then there exists a countable sub-collection of G whose limiting set is K_1.*

Let P be a point of K_1. There exists a sub-collection G_P of G whose limit set exists and contains P. Suppose that there exists a point Q of K_P, the limiting set of G_P, such that Q is not a point of K_1. By Theorem 11, P and

Q are well chained in K_P. But then K_1 is not a maximal well chained sub-set of K. Thus, the supposition that K_P is not a sub-set of K_1 leads to a contradiction. The truth of the theorem follows by Theorem 9.

THEOREM 14. *If G is a collection of connected point sets in a compact space D and K_1 is a maximal connected sub-set of the limiting set of G, there exists a countable sub-collection of G whose limiting set is K_1.*

DEFINITION. A collection G of point sets is said to be a *separable collection* provided that it contains a countable sub-collection H such that if g is an element of G it is either an element of H or the limiting set of some infinite sub-collection of H.

THEOREM 15. *If G is a collection of closed point sets in a separable space D, a sufficient condition that G be a separable collection is that it contain a sub-collection H such that* (1) *H has the same limiting set as G, and* (2) *is upper semi-continuous with respect to G. In particular, if the collection G just mentioned is upper semi-continuous, it is a separable collection.*

The condition mentioned in the theorem is not necessary. Consider, for instance, the collection G whose elements are the graphs in the euclidean plane of $x=0$ and of $y=nx$, where $n=1, 2, 3, \cdots$.

Let K be the common limiting set of G and H. By Theorem 2, G is the sum of a pair of sub-collections G_1 and G_2, and H of a pair of sub-collections G_3 and G_4 such that G_1 and G_3 contain at most a countable number of elements and every element of G_2 and of G_4 respectively is a sub-set of K. Since the space is separable, K is separable and contains a countable sub-set $T=P_1+P_2+P_3+\cdots$ such that every point of it either belongs to T or is a limit point of T. For each positive integer n there exists a countable sub-collection H_n of H having a limit set containing P_n. Let $G_5=\sum H_n$ $(n=1, 2, 3, \cdots, \infty)$. If g is an element of G_2, it contains a countable sub-set $T_g=Q_{1g}+Q_{2g}+Q_{3g}+\cdots$ such that every point of g either belongs to T_g or is a limit point of it. For every positive integral i the point Q_{ig} is a point or a limit point of a sub-sequence of T; it can easily be shown that Q_{ig} belongs to the *limit* set of a sequence M_{ig} belonging to G_5 such that no element of M_{ig} is at a lower distance from Q_{ig} greater than $1/i$. By the note following Theorem 10, g is the limiting set of $\sum M_{ig}$ $(i=1, 2, 3, \cdots, \infty)$. Hence G_1+G_5 is a sub-collection of G of the type required to make G a separable collection.

THEOREM 16. *The aggregate of all closed point sets in a space L may itself be regarded as a space L provided that "the limit of a sequence of elements" in this hyperspace be defined as the minimal closed set in the given space L containing the non-vacuous sequential limiting set of the given sequence.*

If M is a *closed* point set in a space L, the sequential limiting set of the sequence $M, M, M, M, \cdots$, all of whose elements are the point set M, is clearly M. Thus, the third condition in the definition of a space L is satisfied in our hyperspace. If N is a point set in a space L, there exists* a point set K which is the minimal closed point set in that space containing N. From the results just indicated and the definition of a sequential limiting set it follows that our hyperspace satisfies the other requirements of a space L.

THEOREM 17. *The aggregate of all closed point sets in a space S is a space L provided that "the limit of a sequence" is defined as the non-vacuous sequential limiting set of the sequence in the given space S.*

The theorem follows from Theorems 1 and 16.

---

* Cf. Hausdorff II, p. 230.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

RIESZ SUMMABILITY FOR DOUBLE SERIES*

BY

FLORENCE M. MEARS

I. Introduction

Among the definitions for summing simple series is that introduced by M. Riesz† in 1909, and later generalized by him to the following form:

Let $\{\lambda_n\}$ be any sequence of positive numbers, increasing and becoming infinite, with $\lambda_1 \geqq 0$; let p be any real number greater than or equal to zero. Consider the series $\sum_{k=1,\cdots,\infty} c_k$. Write

$$y(s) = s^{-p} \sum_{\lambda_k < s} (s - \lambda_k)^p c_k.$$

If

$$\lim_{s \to \infty} y(s) = C,$$

the series $\sum_{k=1,\cdots,\infty} c_k$ is said to be summable (R, λ, p) to sum C.‡

The purpose of this paper is to extend this definition to one for summability of double series, and to develop some of the properties of this extended definition. We consider the double series

$$\begin{aligned} & u_{11} + u_{12} + u_{13} + \cdots \\ & + u_{21} + u_{22} + u_{23} + \cdots \\ & + \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \end{aligned} \tag{1.1}$$

Let $\{\lambda_n\}$ and $\{\mu_n\}$ be two sequences of real numbers, increasing and becoming infinite, with $\lambda_1 \geqq 0$, $\mu_1 \geqq 0$; let p and r be any real numbers. We define

$$C_{\lambda\mu}^{pr}(s,t) = \sum_{\lambda_k < s} \sum_{\mu_l < t} (s - \lambda_k)^p (t - \mu_l)^r u_{kl}. \tag{1.2}$$

We shall write

$$C_{\lambda\mu}(\sigma,\tau) = C_{\lambda\mu}^{00}(\sigma,\tau) = \sum_{k=1,l=1}^{m,n} u_{kl} \text{ for } \begin{cases} \lambda_m < \sigma \leqq \lambda_{m+1} \\ \mu_n < \tau \leqq \mu_{n+1} \end{cases}.$$

Then for $p > 0$, $r > 0$, we have

---

* Presented to the Society, September 9, 1927; received by the editors in July, 1927.

† Riesz, *Sur la sommation des séries de Dirichlet*, Paris Comptes Rendus, July 5, 1909.

‡ Riesz, *Sur les séries de Dirichlet et les séries entières*, Paris Comptes Rendus, November 22, 1909.

$$C_{\lambda\mu}^{pr}(s,t) = pr \int_0^s \int_0^t C_{\lambda\mu}(\sigma,\tau)(s-\sigma)^{p-1}(t-\tau)^{r-1}\, d\tau d\sigma. \tag{1.3}$$

For $p=0$, $r>0$,

$$C_{\lambda\mu}^{0r}(s,t) = r \int_0^t C_{\lambda\mu}(s,\tau)(t-\tau)^{r-1} d\tau, \tag{1.4}$$

and for $r=0$, $p>0$,

$$C_{\lambda\mu}^{p0}(s,t) = p \int_0^s C_{\lambda\mu}(\sigma,t)(s-\sigma)^{p-1} d\sigma. \tag{1.5}$$

We shall say that the series $\sum_{k,l=1,\cdots,\infty} u_{kl}$ is summable $(R: \lambda, p; \mu, r)$ to sum U, if

$$\lim_{s\to\infty,\, t\to\infty} s^{-p} t^{-r} C_{\lambda\mu}^{pr}(s,t) = U.$$

The present paper includes theorems relative to the regularity and total regularity of the extended definition; it establishes a relation between methods of summation of the same type, λ and μ, when either p or r, or both p and r, are changed; a relation between methods of summation of the same order, p and r, when either λ or μ, or both λ and μ, are changed; certain necessary conditions for the Riesz summability of double series; theorems for the Dirichlet and Cauchy products of double series, corresponding to the theorems of Mertens, Cauchy and Abel for the Cauchy product of simple series; a sufficient condition for the summability of the product of two double series to the correct sum.

The proofs are similar to those of Hardy and Riesz* for the Riesz definition for simple series, with such alterations as might reasonably be expected because of change in dimensionality. Because of this similarity, the proofs are in many cases condensed; where they differ, they are given in more detail.†

---

* Hardy and Riesz, *General Theory of Dirichlet's Series*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 18. We shall refer to this tract by the letters H. R.

† The theorems of the present paper include as special cases results for the summability of bounded double series. Certain of these results for bounded series, obtained by Dr. G. M. Merriman, have been included in his paper *Concerning the summability of double series of a certain type*, Annals of Mathematics, (2), vol. 28, p. 515. We shall refer to this paper by the letter M. The writer's first information regarding Dr. Merriman's work was obtained on reading an abstract of it which appeared in Bulletin of the American Mathematical Society, vol. 33 (1927), p. 407, after the present paper was completely written for publication.

II. Regularity and total regularity

Denote by $\{x_{mn}\}$ the double sequence

$$x_{11},\ x_{12},\ x_{13},\ \cdots$$
$$x_{21},\ x_{22},\ x_{23},\ \cdots$$
$$\cdots\cdots\cdots$$

associated with the double series (1.1). Then

$$x_{mn} = \sum_{k=1,l=1}^{m,n} u_{kl}.$$

A method for evaluating double series is said to be regular if whenever $\{x_{mn}\}$ is a bounded convergent sequence, the transformed sequence converges to the same value. A regular transformation of real elements is said to be totally regular if when applied to a sequence of real elements $\{x_{mn}\}$ which has the following properties,

(a) x_{mn} bounded for each m,

(b) x_{mn} bounded for each n,

(c) $\lim_{m\to\infty,n\to\infty} x_{mn} = +\infty$,

it transforms the sequence into a sequence which has for its limit $+\infty$.

Let

$$s^{-p}t^{-r}C_{\lambda\mu}^{pr}(s,t) = y(s,t) = \sum_{k=1,l=1}^{\infty,\infty} a_{kl}(s,t)x_{kl}. \tag{2.1}$$

In order that a transformation of this form be regular,* it is necessary and sufficient that

(a) $\lim_{s\to\infty,t\to\infty} a_{kl}(s,t) = 0$, for each k and l;

(b) $\lim \sum_{k=1,l=1}^{\infty,\infty} a_{kl}(s,t) = 1$;

(c) $\lim_{s\to\infty,t\to\infty} \sum_{k=1}^{\infty} |a_{kl}(s,t)| = 0$, for each l;

(d) $\lim_{s\to\infty,t\to\infty} \sum_{l=1}^{\infty} |a_{kl}(s,t)| = 0$, for each k;

(e) $\sum_{k=1,l=1}^{\infty,\infty} |a_{kl}(s,t)| < A$, for every pair of values of s and t.

* G. M. Robison, these Transactions, vol. 28 (1926), p. 67.

In order that a transformation of the form (2.1) be totally regular,* it is necessary and sufficient that there exist integers k_0 and l_0 such that $a_{kl}(s, t) \geqq 0$, when $k > k_0$, $l > l_0$, for all s and t.

We state without proof the following theorems.

THEOREM I. *$(R: \lambda, p; \mu, r)$ is regular for $p \geqq 0$, $r \geqq 0$.*

THEOREM II. *$(R: \lambda, p; \mu, r)$ is totally regular for $p \geqq 0$, $r \geqq 0$.*

III. LEMMAS

We shall derive eight lemmas, which are necessary for the proofs of the subsequent theorems.

We make use of the symbols o, O_x, O_y, and O, which we define as follows.

If ϕ is a positive function, we shall write

$$f(x, y) = o(\phi(x, y)),$$

if $\lim (f/\phi) = 0$, as $x \to \infty$, $y \to \infty$, independently.

If ϕ is a positive function, and if to any pair of constants, α_1, β_1, where $0 \leqq \alpha_1 \leqq \beta_1$, there corresponds a constant M_1, such that

$$\frac{|f|}{\phi} < M_1, \text{ for } \alpha_1 \leqq x \leqq \beta_1, \text{ and for all } y,$$

we write

$$f(x, y) = O_x(\phi(x, y)).$$

If ϕ is a positive function, and if to any pair of constants, α_2, β_2, where $0 \leqq \alpha_2 \leqq \beta_2$, there corresponds a constant M_2, such that

$$\frac{|f|}{\phi} < M_2, \text{ for } \alpha_2 \leqq y \leqq \beta_2, \text{ and for all } x,$$

we write

$$f(x, y) = O_y(\phi(x, y)).$$

If ϕ is a positive function, and if there exists a constant M_3, such that

$$\frac{|f|}{\phi} < M_3, \text{ for all positive values of } x \text{ and } y,$$

we write

$$f(x, y) = O(\phi(x, y)).$$

It is clear that if $f(x, y) = O_x(\phi(x, y))$, $f(x, y) = O_y(\phi(x, y))$ and $f(x,y) = o(\phi(x, y))$, then $f(x, y) = O(\phi(x, y))$.

---

* Robison, loc. cit., p. 70.

LEMMA 1. *Let $\psi(x)$ be a continuous function such that*

$$\psi(x) \sim Ax^{\alpha} \text{ as } x \to \infty \text{ , where } \alpha > -1 ;$$

let $\phi(x, y)$ be a function of x and y, and let

$$\chi(x,y) = \int_0^x \phi(u,y)\psi(x-u)du.$$

Then

(a) *if* $\phi(x,y) = O_y(x^{\beta})$, *and* $\beta > -1$,

$$\chi(x,y) = O_y(x^{\alpha+\beta+1}) ;$$

(b) *if* $\phi(x,y) = O_x(y^{\gamma})$,

$$\chi(x,y) = O_x(y^{\gamma}) ;$$

(c) *if* $\phi(x,y) \sim Bx^{\beta}y^{\gamma}$ *as* $x \to \infty$, $y \to \infty$, *and if* (b) *is satisfied,*

$$\chi(x,y) \sim AB\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}x^{\alpha+\beta+1}y^{\gamma}.$$

(a) We can write

$$\psi(u) = Au^{\alpha} + \psi_1(u), \text{ where } \psi_1(u) = o(u^{\alpha}).$$

By hypothesis, constants y_1 and M_1 exist such that for $0 \leqq y \leqq y_1$,

$$\begin{aligned} |\chi(x,y)| &\leqq \int_0^x \frac{|\phi(u,y)|}{u^{\beta}} |A(x-u)^{\alpha} + \psi_1(x-u)|\, u^{\beta}du \\ &< M_1 \int_0^x |A(x-u)^{\alpha} + \psi_1(x-u)|\, u^{\beta}du \\ &= M_1\left\{ |A| \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} x^{\alpha+\beta+1} + \int_0^x |\psi_1(u')(x-u')^{\beta}|\, du' \right\}. \end{aligned}$$

Given $\epsilon > 0$, we can choose ξ so that

$$\begin{aligned} \int_0^x |\psi_1(u')(x-u')^{\beta}|\, du' &< M_2 \int_0^{\xi} (u')^{\alpha}(x-u')^{\beta}du' \\ &\quad + \epsilon \int_{\xi}^x (u')^{\alpha}(x-u')^{\beta}du' = J_1 + J_2, \end{aligned}$$

where M_2 is a number depending only on ξ. Since $(x-u')^{\beta}$ is increasing throughout the interval $(0, \xi)$, or decreasing throughout this interval, by the second law of the mean for integrals, and for $0 < \xi_1 < \xi$,

$$J_1 = M_2\left[(x-\xi)^\beta \int_{\xi_1}^{\xi} (u')^\alpha du' + (x)^\beta \int_0^{\xi_1} (u')^\alpha du'\right] = O(x^\beta) = o(x^{\alpha+\beta+1}).$$

We have

$$J_2 \leqq \epsilon \int_0^x (u')^\alpha (x-u')^\beta du' = o(x^{\alpha+\beta+1}).$$

Therefore $|\chi(x, y)|$ is of the desired form.

(b) For $0 \leqq x \leqq x_1$,

$$|\chi(x,y)| \leqq \int_0^x \left|\frac{\phi(u,y)\psi(x-u)}{y^\gamma (x-u)^\alpha}\right| (x-u)^\alpha y^\gamma du$$

$$= y^\gamma M_3 \int_0^x (x-u)^\alpha du = O_x(y^\gamma),$$

where x_1 and M_3 are constants.

(c) Write $\psi(u) = Au^\alpha + \psi_1(u)$, where $\psi_1(u) = o(u^\alpha)$, and $\phi(u, y) = Bu^\beta y^\gamma + \phi_1(u, y)$, where $\phi_1 = o(u^\beta y^\gamma)$. Substituting in $\chi(x, y)$ we obtain the sum of four integrals. We have

$$(3.11) \qquad AB^\gamma \int_0^x u^\beta (x-u)^\alpha du = AB\, y^\gamma \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} x^{\alpha+\beta+1}.$$

It is necessary to prove that the other three integrals are of the form $o(x^{\alpha+\beta+1}y^\gamma)$.

Given $\epsilon > 0$, we can choose ξ and η so that

$$|\phi_1(u,y)| < \epsilon u^\beta y^\gamma \quad \text{for } u \geqq \xi,\ y \geqq \eta.$$

Assume $y \geqq \eta$.

We have

$$\int_0^x |A\phi_1(u,y)(x-u)^\alpha|\, du$$

$$= \int_0^\xi |A\phi_1(u,y)(x-u)^\alpha|\, du + \int_\xi^x |A\phi_1(u,y)(x-u)^\alpha|\, du$$

$$< |A| \left\{ M y^\gamma \int_0^\xi (x-u)^\alpha du + \epsilon y^\gamma \int_\xi^x u^\beta (x-u)^\alpha du \right\}$$

where M is a constant.

Applying the second law of the mean for integrals to $\int_0^\xi y^\gamma (x-u)^\alpha du$, we have

$$(3.12)\qquad \int_0^x |A\phi_1(u,y)(x-u)^\alpha|\,du = O(x^\alpha y^\gamma) + o(x^{\alpha+\beta+1}y^\gamma) = o(x^{\alpha+\beta+1}y^\gamma),$$

which is the desired form.

For the same ξ and η,

$$\begin{aligned}
&\int_0^x |\,\phi_1(u,y)\psi_1(x-u)\,|\,du \\
&= \int_0^\xi |\phi_1(u,y)\psi_1(x-u)|\,du + \int_\xi^x |\,\phi_1(u,y)\psi_1(x-u)\,|\,du \\
&< My^\gamma \int_0^\xi |\,\psi_1(x-u)\,|\,du + \epsilon y^\gamma \int_\xi^x u^\beta |\,\psi_1(x-u)\,|\,du \\
&= My^\gamma M_4\xi + \epsilon y^\gamma \int_0^{x-\xi} |\,(x-u')^\beta\psi_1(u')\,|\,du',
\end{aligned}$$

where M_4 depends only on ξ.

By the proof for J_1 and J_2 in (a),

$$\epsilon y^\gamma \int_0^{x-\xi} |\,(x-u')^\beta\psi_1(u')\,|\,du' = o(x^{\alpha+\beta+1}y^\gamma)$$

and therefore

$$(3.13)\qquad \int_0^x |\,\phi_1(u,y)\psi_1(x-u)\,|\,du = o(x^{\alpha+\beta+1}y^\gamma).$$

From the proof used for J_1 and J_2,

$$(3.14)\qquad \int_0^x |\,Bu^\beta y^\gamma\psi_1(x-u)\,|\,du = o(x^{\alpha+\beta+1}y^\gamma).$$

From the results for (3.11), (3.12), (3.13) and (3.14) it follows that $\chi(x,y)$ is of the desired form.

LEMMA 2. *Let $\phi(x,y)$ and $\psi(x,y)$ be functions of x and y, and let*

$$\chi(x,y) = \int_0^x \int_0^y \psi(u,v)\phi(x-u,y-v)\,dv\,du.$$

Then

(a) *if $\psi(x,y)=O_y(x^\alpha)$ and $\phi(x,y)=O_y(x^\gamma)$ where α and γ are >-1,*

$$\chi(x,y) = O_y(x^{\alpha+\gamma+1});$$

(b) *if $\psi(x,y)=O_x(y^\beta)$ and $\phi(x,y)=O_x(y^\delta)$ where β and δ are >-1,*

$$\chi(x,y) = O_x(x^{\beta+\delta+1});$$

(c)* *if* $\psi(x, y) \sim Ax^{\alpha}y^{\beta}$† *and* $\phi(x, y) \sim Bx^{\gamma}y^{\delta}$ *as* $x \to \infty$, $y \to \infty$, *and if* (a) *and* (b) *are satisfied,*

$$\chi(x,y) \sim AB \frac{\Gamma(\alpha+1)\Gamma(\beta+1)\Gamma(\gamma+1)\Gamma(\delta+1)}{\Gamma(\alpha+\gamma+2)\Gamma(\beta+\delta+2)} x^{\alpha+\gamma+1}y^{\beta+\delta+1}.$$

(a) We can find a constant M, such that when $0 \leqq v \leqq y_1$,

$$\left|\frac{\psi(u,v)}{u^{\alpha}}\right| < M \text{ and } \left|\frac{\phi(x-u, y-v)}{(x-u)^{\gamma}}\right| < M.$$

Assuming $0 \leqq y \leqq y_1$, we have

$$|\chi(x,y)| < M^2 \int_0^x \int_0^y u^{\alpha}(x-u)^{\gamma} dv\, du$$

$$\leqq M^2 \frac{\Gamma(\alpha+1)\Gamma(\gamma+1)}{\Gamma(\alpha+\gamma+2)} x^{\alpha+\gamma+1}y_1 = O_y(x^{\alpha+\gamma+1}).$$

(b) The proof of (b) is similar to that of (a).

(c) Using the method used in M., p. 519, we write $\chi(x, y)$ as the sum of four integrals, the first of which is of the desired form.

We write

$$\int_0^x \int_0^y Au^{\alpha}v^{\beta}\phi_1(x-u, y-v)dv\, du \tag{3.21}$$

$$= \int_0^x \int_0^y A(x-u')^{\alpha}(y-v')^{\beta}\phi_1(u', v')dv'du'$$

in the form

$$\left|\int_0^x \int_0^y\right| \leqq \int_0^{\xi} \int_0^y + \int_0^x \int_0^{\eta} + \int_{\xi}^x \int_{\eta}^y,$$

where the integrands on the right hand side of the inequality are in absolute value. Given $\epsilon > 0$, we choose ξ and η so that

$$\phi_1(u', v') < \epsilon(u')^{\gamma}(v')^{\delta}, \text{ for } \xi \leqq u', \eta \leqq v',$$

and by methods similar to those used in Lemma 1, we can show that the integral of (3.21) is of the desired form.

Similarly, we can show that

$$\int_0^x \int_0^y B(x-u)^{\gamma}(y-v)^{\delta}\psi_1(u,v)dv\, du = o(x^{\alpha+\gamma+1}y^{\beta+\delta+1}). \tag{3.22}$$

---

* Case (c) corresponds to M., p. 519, Lemma 4. The conditions (a) and (c) are replaced here by the less stringent conditions (a) and (b).

† By $\psi(x, y) \sim Ax^{\alpha}y^{\beta}$ we mean that $\psi/(x^{\alpha}y^{\beta}) \to A$ as $x \to \infty$, $y \to \infty$, independently.

It remains to prove that

$$\int_0^x \int_0^y \psi_1(u,v)\phi_1(x-u,y-v)dv\,du = o(x^{\alpha+\gamma+1}y^{\beta+\delta+1}). \tag{3.23}$$

Given $\epsilon>0$, we can choose ξ and η so that

$$\psi_1(u,v) < \epsilon u^\alpha v^\beta, \text{ for } \xi \leqq u,\ \eta \leqq v.$$

We express the absolute value of the integral as the sum of three integrals, J_1', J_2', and J_3', as in (3.21).

We have

$$\begin{aligned} J_3' &< \epsilon \int_\xi^x \int_\eta^y u^\alpha v^\beta \phi_1(x-u,y-v)dv\,du \\ &= \epsilon \int_0^{x-\xi} \int_0^{y-\eta} (x-u')^\alpha (y-v')^\beta \phi_1(u',v')dv'du'. \end{aligned}$$

We can choose ξ_1 and η_1 such that

$$\phi_1(u',v') < \epsilon(u')^\gamma (v')^\delta, \text{ for } \xi_1 \leqq u',\ \eta_1 \leqq v'.$$

We express J_3' as the sum of three integrals as in (3.21) and by methods similar to those used in Lemma 1, we can show that J_1', J_2', and J_3', and hence (3.23), are of the form $o(x^{\alpha+\gamma+1}y^{\beta+\delta+1})$.

From the results for (3.21), (3.22), and (3.23), it follows that $\chi(x, y)$ is of the desired form.

LEMMA 3. *If* $p \geqq 0$, $p'>0$, $r \geqq 0$,

$$C_{\lambda\mu}^{p+p',r}(s,t) = \frac{\Gamma(p+p'+1)}{\Gamma(p+1)\Gamma(p')} \int_0^s C_{\lambda\mu}^{pr}(u,t)(s-u)^{p'-1}du.$$

We have a similar result if $r \geqq 0$, $r'>0$, $p \geqq 0$.

We omit the proof, because of its similarity to that of H. R., Lemma 6.

LEMMA 4. *If* $p \geqq 0$, $p'<1$, $p' \leqq p$ *and* $r \geqq 0$,

$$C_{\lambda\mu}^{p-p',r}(s,t) = \frac{\Gamma(p-p'+1)}{\Gamma(p+1)\Gamma(1-p')} \int_0^s \frac{\partial C_{\lambda\mu}^{pr}(u,t)}{\partial u}(s-u)^{-p'}du.$$

We have a similar result if $r \geqq 0$, $r' \leqq r$, $r'<1$ and $p \geqq 0$. The proof is obtained from the preceding lemma by a method analogous to that of M., Lemma 1, part 4.

LEMMA 5. *If u_{kl} is real, $0 \leqq \eta \leqq t$, $p' \geqq 0$, $r' \geqq 0$, $0 < r \leqq 1$, then*

$$\frac{\Gamma(r+r'+1)}{\Gamma(r'+1)\Gamma(r)} \left| \int_0^\eta C_{\lambda\mu}^{p'r'}(\sigma,\tau)(t-\tau)^{r-1}d\tau \right| \leqq \max_{0\leqq\tau\leqq\eta} \left| C_{\lambda\mu}^{p',r+r'}(\sigma,\tau) \right| .$$

We have a similar result for $0 \leqq \xi \leqq s$, $r' \geqq 0$, $p' \geqq 0$, $0 < p \leqq 1$. The proof is analogous to that of H.R., Lemma 8, and M., Lemma 2.

LEMMA 6. *If $p \geqq 0$, $r \geqq 0$, $p' < 1$, $r' < 1$, $p' \leqq p$ and $r' \leqq r$, then*

$$C_{\lambda\mu}^{p-p',r-r'}(s,t) = \frac{\Gamma(p-p'+1)\Gamma(r-r'+1)}{\Gamma(p+1)\Gamma(1-p')\Gamma(r+1)\Gamma(1-r')} \cdot \int_0^s \int_0^t \frac{\partial^2 C_{\lambda\mu}^{pr}(u,v)}{\partial v \partial u}(s-u)^{-p'}(t-v)^{-r'}dv\,du.$$

The proof is obtained by applying Lemma 4 twice.

LEMMA 7. *If u_{kl} is real, $0 \leqq \xi \leqq s$, $0 \leqq \eta \leqq t$, $p' \geqq 0$, $r' \geqq 0$, $0 < p \leqq 1$ and $0 < r \leqq 1$, then*

$$\frac{\Gamma(p+p'+1)\Gamma(r+r'+1)}{\Gamma(p'+1)\Gamma(p)\Gamma(r'+1)\Gamma(r)} \left| \int_0^\xi \int_0^\eta C_{\lambda\mu}^{p'r'}(\sigma,\tau)(s-\sigma)^{p-1}(t-\tau)^{r-1}d\tau\,d\sigma \right| \leqq \max_{\substack{0\leqq\sigma\leqq\xi \\ 0\leqq\tau\leqq\eta}} \left| C_{\lambda\mu}^{p+p',r+r'}(\sigma,\tau) \right| .$$

The proof is analogous to the proofs referred to in Lemma 5.

LEMMA 8. *If ϕ is a positive function of x such that*

$$\int^x \phi dx \to \infty \ , \ as \ x \to \infty \ ,$$

and if f is a function of x and y, such that

$$f = o(\phi), \ and \ f = O_x(1),$$

then

$$\int^x o(\phi)dx = o\left\{ \int^x \phi\, dx \right\}.$$

Since $f = o(\phi)$, for any $\epsilon > 0$ we can find ξ and η such that

$$|f| < \epsilon\phi, \ \text{for } x \geqq \xi \text{ and } y \geqq \eta.$$

Since $f = O_x(1)$, there exists some constant M, such that

$$|f| < M, \ \text{for } x < \xi.$$

Assuming $y \geqq \eta$, we have

$$\left| \int_0^x f(x,y)dx \right| \leqq \left| \int_a^\xi f(x,y)dx \right| + \left| \int_\xi^x f(x,y)dx \right|$$

$$< M(\xi - a) + \epsilon \int_\xi^x \phi dx \leqq M(\xi - a) + \epsilon \int_a^x \phi dx.$$

Therefore

$$\frac{\left| \int_a^x f(x,y)dx \right|}{\int_a^x \phi(x)dx} < \frac{M(\xi - a)}{\int_a^x \phi(x)dx} + \epsilon.$$

Since $\int_a^x \phi dx$ diverges, we have the desired result.

IV. Theorems of consistency

THEOREM III. *If the series $\sum u_{mn}$ is bounded and summable $(R: \lambda, p; \mu, r)$ to sum U, then $\sum u_{mn}$ is summable $(R: \lambda, p'; \mu, r)$ to sum U, provided $p' \geqq p \geqq 0$, $r \geqq 0$. Both $s^{-p}t^{-r}C_{\lambda\mu}^{pr}(s,t)$ and $s^{-p'}t^{-r}C_{\lambda\mu}^{p'r}(s,t)$ are bounded.*

It has been proved that a bounded sequence is transformed by a regular transformation into a bounded sequence;* therefore Theorem III is a special case of the following more general theorem.

THEOREM III′. *If $\sum u_{mn}$ is summable $(R: \lambda, p; \mu, r)$ to sum U, and if $C_{\lambda\mu}^{pr}(s,t) = O_s(t^r)$, then $\sum u_{mn}$ is summable $(R: \lambda, p'; \mu, r)$ to sum U, and $C_{\lambda\mu}^{p'r}(s,t) = O_s(t^r)$, provided $p' \geqq p \geqq 0$, and $r \geqq 0$. If $C_{\lambda\mu}^{pr}(s,t) = O(s^p t^r)$, then $C_{\lambda\mu}^{p'r}(s,t) = O(s^{p'}t^r)$.*

The theorem is obviously true for $p' = p \geqq 0$. We assume that $p' = p + p''$, where $p'' > 0$; from Lemma 3, we have

$$(4.1) \qquad C_{\lambda\mu}^{p+p'',r}(s,t) = \frac{\Gamma(p+p''+1)}{\Gamma(p+1)\Gamma(p'')} \int_0^s C_{\lambda\mu}^{pr}(u,t)(s-u)^{p''-1}du.$$

Since by hypothesis $C_{\lambda\mu}^{pr}(s,t) \sim Us^p t^r$ as $s \to \infty$, $t \to \infty$, and since $C_{\lambda\mu}^{pr}(s,t) = O_s(t^r)$, we have, by Lemma 1(c),

$$\int_0^s C_{\lambda\mu}^{pr}(u,t)(s-u)^{p''-1}du \sim U \frac{\Gamma(p'')\Gamma(p+1)}{\Gamma(p+p''+1)} s^{p+p''}t^r,$$

and by Lemma 1(b)

---

* Robison, loc. cit., p. 67.

$$\int_0^s C_{\lambda\mu}^{pr}(u,t)(s-u)^{p''-1}\,du = O_s(t^r).$$

If $C_{\lambda\mu}^{pr}(s,t) = O(s^p t^r)$, we have, by Lemma 1(a),

$$\int_0^s C_{\lambda\mu}^{pr}(u,t)(s-u)^{p''-1}\,du = O_t(s^{p+p''}).$$

Using these results in (4.1), we have the desired theorem.

The analogous theorem for $p \geqq 0$ and $r' \geqq r \geqq 0$ may be proved in the same way. If we change p to p' under the condition of Theorem III, and then change r to r' under the conditions of the analogous theorem, we obtain the following theorem.

THEOREM IV. *If the series $\sum u_{mn}$ is bounded and summable $(R: \lambda, p; \mu, r)$ to sum U, then $\sum u_{mn}$ is summable $(R: \lambda, p'; \mu, r')$ to sum U, provided $p' \geqq p \geqq 0$ and $r' \geqq r \geqq 0$. Both $s^{-p}t^{-r}C_{\lambda\mu}^{pr}(s,t)$ and $s^{-p'}t^{-r'}C_{\lambda\mu}^{p'r'}(s,t)$ are bounded.**

This theorem is a special case of the following theorem.

THEOREM IV′. *If $\sum u_{mn}$ is summable $(R: \lambda, p; \mu, r)$ to sum U, and if $C_{\lambda\mu}^{pr}(s,t) = O(s^p t^r)$, then $\sum u_{mn}$ is summable $(R: \lambda, p'; \mu, r')$ to sum U, and $C_{\lambda\mu}^{p'r'}(s,t) = O(s^{p'}t^{r'})$, provided $p' \geqq p \geqq 0$, $r' \geqq r \geqq 0$.*

This group of theorems is a generalization for double series of the first theorem of consistency, established by Hardy and Riesz† for simple series. For the generalization for double series of the second theorem of consistency‡ we require the following lemmas,§ which we state without proof.

LEMMA 9. *Any logarithmico-exponential function (or, in short, L-function) $\mu(\lambda)$ is continuous, of constant sign, and monotonic from a certain value of λ onwards; and the same is true of any of its derivatives.*

---

* Cf. M., p. 521, Theorem I. This theorem is true for bounded series. That the proof, even with this restriction, does not follow from the statement $C^{r,s}(\sigma,\tau_1) \approx C_1\sigma^r, C^{r,s}(\sigma_1,\tau) \approx C_2\tau^s$, is shown by the series

$$c_{1n} = (-1)^{n-1}; \quad c_{mn} = (-1)^n \frac{1}{2^{m-1}}, \text{ for } m \neq 1,$$

which is bounded and convergent to zero. If $\lambda_n = 2^{n-1}$, $\lambda_m = 2^{m-1}$, the series is C^{11} summable, but $C^{11}(\sigma_1,\tau) \approx C_2\tau$ is not true.

† H. R., p. 29.

‡ H. R., p. 30; Hardy, *The second theorem of consistency for summable series*, Proceedings of the London Mathematical Society, (2), vol. 15 (1915), pp. 72–88.

§ Hardy, *The second theorem* etc., p. 75.

LEMMA 10. *If $\mu\to\infty$, and a number Δ exists such that*

$$\mu = O(\lambda^{\Delta}),$$

then

$$\mu^{(r)} = O\left(\frac{\mu}{\lambda^r}\right),$$

$\mu^{(r)}$ denoting the rth derivative of $\mu(\lambda)$.

LEMMA 11. *If μ satisfies the conditions of Lemma 10, and if ν lies between two positive numbers g and G, then positive numbers h and H exist, such that*

$$h \leqq \frac{\mu(\lambda\nu)}{\mu(\lambda)} \leqq H.$$

The same result holds for decreasing functions which decrease less rapidly than $\lambda^{-\Delta}$ for some value of Δ.

The first theorem of the group corresponding to the second theorem of consistency is as follows.

THEOREM V. *If the series $\sum u_{kl}$ is bounded and summable $(R: \lambda, p; \mu, r)$ to sum U, and if η is an L-function of λ, such that*

$$\eta = O(\lambda^{\Delta}),$$

where Δ is a constant, then $\sum u_{kl}$ is summable $(R: \eta, p; \mu, r)$ to sum U, provided $p \geqq 0$, $r \geqq 0$. Both $s^{-p}t^{-r}C_{\lambda\mu}^{pr}(s, t)$ and $s^{-p}t^{-r}C_{\eta\mu}^{pr}(s, t)$ are bounded.

This is a special case of the following theorem.

THEOREM V′. *If the series $\sum u_{kl}$ is summable $(R: \lambda, p; \mu, r)$ to sum U and if $C_{\lambda\mu}^{pr}(s, t) = O_s(t^r)$; if η is an L-function of λ, such that*

$$\eta = O(\lambda^{\Delta}),$$

where Δ is a constant, then $\sum u_{kl}$ is summable $(R: \eta, p; \mu, r)$ to sum U, and $C_{\eta\mu}^{pr}(s, t) = O_s(t^r)$, provided $p \geqq 0$ and $r \geqq 0$. If $C_{\lambda\mu}^{pr}(s, t) = O(s^p t^r)$, then $C_{\eta\mu}^{pr}(s, t) = O(s^p t^r)$.

We shall assume that u_{kl} is real and that $U = 0$. If u_{kl} is complex, we can consider the real and imaginary parts of the series separately. If $U \neq 0$, we prove the theorem for the double series $\sum u'_{kl}$, where

$$u'_{kl} = \begin{cases} u_{11} - U, & \text{for } k = l = 1, \\ u_{kl}, & \text{for all other } k \text{ and } l. \end{cases}$$

We complete the proof by adding the convergent double series $\sum u''_{kl}$, where

$$u''_{kl} = \begin{cases} U, & \text{for } k = l = 1, \\ 0, & \text{for all other } k \text{ and } l. \end{cases}$$

Case 1. $p=0$, $r\geqq 0$.

In $C^{0r}_{\eta\mu}(s, w)$, let $s=\eta(v)$. Since $\eta_m=\eta(\lambda_m)<\eta(v)\leqq\eta_{m+1}$ when $\lambda_m<v\leqq\lambda_{m+1}$, it follows immediately from the definition of $C^{00}_{\lambda\mu}(\sigma, \tau)$ that

$$C^{00}_{\eta\mu}[\eta(v), w] = C^{00}_{\lambda\mu}(v, w)$$

and that therefore

$$C^{0r}_{\eta\mu}[\eta(v), w] = C^{0r}_{\lambda\mu}(v, w).$$

Hence the theorem is true under the conditions of Case 1.

From Lemma 3, for $p>0$, $r\geqq 0$, we have

$$C^{pr}_{\lambda\mu}(v, w) = p\int_{\lambda_1}^{v} C^{0r}_{\lambda\mu}(v, w)(v-\nu)^{p-1}d\nu = o(v^p w^r). \tag{4.21}$$

We wish to prove that

$$C^{pr}_{\eta\mu}(s, w) = p\int_{\eta_1}^{s} C^{0r}_{\eta\mu}(\sigma, w)(s-\sigma)^{p-1}d\sigma = o(s^p w^r). \tag{4.22}$$

Let $s=\eta(v)$ and $\sigma=\eta(\nu)=\eta$.

Since

$$C^{0r}_{\eta\mu}[\eta(\nu), w] = C^{0r}_{\lambda\mu}(\nu, w),$$

we may write (4.22) in the form

$$p\int_{\lambda_1}^{v} C^{0r}_{\lambda\mu}(\nu, w)(s-\eta)^{p-1}\eta' d\nu = o(s^p w^r). \tag{4.23}$$

Case 2. Assume p a positive integer and $r\geqq 0$.

By repeated applications of (3.4), we have

$$C^{0r}_{\lambda\mu}(\nu, w) = \frac{1}{p!}\left(\frac{\partial}{\partial\nu}\right)^p C^{pr}_{\lambda\mu}(\nu, w).$$

Substituting this expression in (4.23), we have a constant multiple of

$$J = \int_{\lambda_1}^{v}\left(\frac{\partial}{\partial\nu}\right)^p C^{pr}_{\lambda\mu}(\nu, w)(s-\eta)^{p-1}\eta' d\nu.$$

Integrating J by parts p times, we obtain

$$J = (-1)^{p-1}\left[C_{\lambda\mu}^{pr}(\nu, w)\left(\frac{\partial}{\partial\nu}\right)^{p}(s-\eta)^{p}\right]_{\nu=\lambda_1}^{\nu=v}$$
$$+(-1)^{p}\int_{\lambda_1}^{v} C_{\lambda\mu}^{pr}(\nu, w)\left(\frac{\partial}{\partial\nu}\right)^{p+1}(s-\eta)^{p}\,d\nu$$
$$=(-1)^{p-1}C_{\lambda\mu}^{pr}(v, w)\left[\left(\frac{\partial}{\partial\nu}\right)^{p}(s-\eta)^{p}\right]_{\nu=v}$$
$$+(-1)^{p}\int_{\lambda_1}^{v} C_{\lambda\mu}^{pr}(\nu, w)\left(\frac{\partial}{\partial\nu}\right)^{p+1}(s-\eta)^{p}\,d\nu$$
$$=J_1+J_2.$$

Since

$$\left[\left(\frac{\partial}{\partial\nu}\right)^{p}\right]_{\nu=v}(-1)^{p}p!(s')^{p},$$

where

$$s'=\frac{\partial}{\partial\nu}\eta(\nu),\ \text{when } \nu=v,$$

we have

$$J_1=-p!C_{\lambda\mu}^{pr}(v, w)(s')^{p}=-p!\frac{C_{\lambda\mu}^{pr}(v, w)}{v^{p}w^{r}}\left(\frac{vs'}{s}\right)^{p}s^{p}w^{r} \tag{4.31}$$
$$=o(1)O(1)s^{p}w^{r}=o(s^{p}w^{r}),$$

by Lemma 10.

We can easily prove that

$$\left(\frac{\partial}{\partial\nu}\right)^{p+1}(s-\eta)^{p}=\sum As^{p-\alpha}\eta^{\beta_0}(\eta')^{\beta_1}(\eta'')^{\beta_2}\cdots,$$

where the A's are constants, and

$$0<\beta_0+\beta_1+\beta_2+\cdots=\alpha\leqq p, \tag{4.32}$$
$$\beta_1+2\beta_2+3\beta_3+\cdots=p+1.$$

Therefore J_2 is the sum of constant multiples of integrals of the form

$$s^{p-\alpha}\int_{\lambda_1}^{v} C_{\lambda\mu}^{pr}(\nu, w)\eta^{\beta_0}(\eta')^{\beta_1}(\eta'')^{\beta_2}\cdots d\nu. \tag{4.33}$$

Since $C_{\lambda\mu}^{pr}(\nu, w)=o(\nu^{p}w^{r})$, and since, by Lemma 10,

$$\eta^{(m)}=O\left(\frac{\eta}{\nu^{m}}\right),$$

(4.33) is of the form

$$s^{p-\alpha}\int_{\lambda_1}^{v} o(\nu^p w^r)\eta^{\beta_0}\eta'\left[O\left(\frac{\eta}{\nu}\right)\right]^{\beta_1-1}\left[O\left(\frac{\eta}{\nu^2}\right)\right]^{\beta_2}\cdots$$

$$(4.34)\qquad = s^{p-\alpha}\int_{\lambda_1}^{v} o\left\{w^r\nu^{p-\beta_1-2\beta_2+\cdots+1}\eta^{\beta_0+\beta_1+\cdots-1}\eta'\right\}d\nu$$

$$= s^{p-\alpha}w^r\int_{\lambda_1}^{v} o\left\{\eta^{\alpha-1}\eta'\right\}d\nu.$$

The conditions of Lemma 8 are satisfied, since

$$\int^{\nu}\eta^{\alpha-1}\eta'd\nu = \frac{\eta^\alpha}{\alpha}\to\infty\,,\quad \text{as } \nu\to\infty\,,$$

and since

$$\frac{C^{pr}_{\lambda\mu}(\nu,w)}{w^r}\,\eta^{\beta_0}(\eta')^{\beta_1}(\eta'')^{\beta_2}\cdots = O_s(1).$$

Therefore we may write (4.34) as

$$(4.35)\qquad s^{p-\alpha}w^r o\left\{\int_{\lambda_1}^{v}\eta^{\alpha-1}\eta'd\nu\right\} = (s^p w^r).$$

From (4.31) and (4.35) we have

$$J = o(s^p w^r),$$

which proves that the series is summable $(R:\ \eta, p; \mu, r)$ under the conditions of Case 2. To complete the proof for Case 2, we must show that under the same conditions, $C^{pr}_{\eta\mu}(s,\ t)=O_s(t^r)$, $[C^{pr}_{\eta\mu}=O(s^p t^r)]$, provided $C^{pr}_{\lambda\mu}=O_s(t^r)$, $[C^{pr}_{\lambda\mu}=O(s^p t^r)]$.

Assume s bounded; then v is bounded. Let $w\to\infty$. From (4.31) we have

$$(4.41)\qquad J_1 = -\,p!\,\frac{C^{pr}_{\lambda\mu}(v,w)}{v^p w^r}O(1)s^p w^r$$

$$= O_s(1)O(1)w^r = O_s(w^r).$$

If we assume w bounded, while $s\to\infty$, and hence $\nu\to\infty$, we have

$$(4.42)\qquad J_1 = O_w(1)O(1)s^p = O_w(s^p).$$

The integral of (4.33) may be written

$$s^{p-\alpha}\int_{\lambda_1}^{v} C^{pr}_{\lambda\mu}(\nu,w)\eta^{\beta_0}\eta'\left[O\left(\frac{\eta}{\nu}\right)\right]^{\beta_1-1}\left[O\left(\frac{\eta}{\nu^2}\right)\right]^{\beta_2}\cdots.$$

If $w\to\infty$, while s and v are bounded, this integral is of the form

$$s^{p-\alpha}O_s(w^r)\int_{\lambda_1}^{v}\nu^{-p}\eta^{\alpha-1}\eta'd\nu = O_s(w^r). \tag{4.43}$$

If we assume w bounded, while s and $v\to\infty$, we have

$$\begin{aligned} &s^{p-\alpha}\int_{\lambda_1}^{v}O_w(1)\nu^{p-\beta_1-2\beta_2-\cdots+1}\eta^{\alpha-1}\eta'd\nu \\ &= s^{p-\alpha}O_w(1)\int_{\lambda_1}^{v}\eta^{\alpha-1}\eta'd\nu = O_w(s^p). \end{aligned} \tag{4.44}$$

From (4.41), (4.42), (4.43), and (4.44) we see that the theorem is true under the conditions of Case 2.

To complete the theorem, we must consider the cases $0<p<1$, $r\geqq 0$, and $p>1$ and not integral, $r\geqq 0$. The proof that the series is summable $(R\colon \eta, p; \mu, r)$ under the conditions of Case 2 is the same as the proof of the corresponding case of the second theorem of consistency* for simple series, if, in the latter proof, we replace $C_{\lambda^\kappa}(\sigma)$ by $C^{pr}_{\lambda\mu}(\nu, w)$ and $o(\zeta^\kappa)$ by $o(s^pw^r)$, letting $p=\kappa$, $\lambda=\lambda$, and making appropriate changes in ν and s for changes in σ and ζ. The proof that the series is summable under the conditions of the remaining cases may be obtained in the same way from the corresponding cases of the theorem for simple series; the extended proof requires Lemmas 5 and 8, together with the group of lemmas 9, 10 and 11. As in Case 2, the fact that $C^{pr}_{\eta\mu}(s, t)=O_s(t^r)$, $[C^{pr}_{\eta\mu}=O(s^pt^r)]$, may be deduced from the equations used in proving the summability.

THEOREM VI. *If the series $\sum u_{kl}$ is summable $(R\colon \lambda, 0; \mu, r)$ to sum U, and if η is an increasing function of λ, $r\geqq 0$, then $\sum u_{kl}$ is summable $(R\colon \eta, 0; \mu, r)$. If $C^{0r}_{\lambda\mu}(s, t)$ is bounded for s bounded, or t bounded, or both bounded, $C^{0r}_{\eta\mu}(s, t)$ is bounded similarly.*

This theorem follows immediately from the proof of Case 1, Theorem V′.

If we hold λ fixed, and let $\theta=O(\mu^\Delta)$, and if $p\geqq 0$, $r\geqq 0$, we have a theorem analogous to Theorem V. The theorems corresponding to Theorems V′ and VI are also true. If we now change λ to η under the conditions of Theorem V, and change μ to θ under the conditions of the analogous theorem, we have the following theorem of consistency.

THEOREM VII. *If the series $\sum u_{kl}$ is bounded and summable $(R\colon \lambda, p; \mu, r)$ to sum U, and if η is an L-function of λ, and θ an L-function of μ, such that*

$$\eta = O(\lambda^{\Delta_1}) \text{ and } \theta = O(\mu^{\Delta_2}).$$

* Hardy, loc. cit., pp. 77. 78.

where Δ_1 and Δ_2 are constants, then $\sum u_{kl}$ is summable $(R: \eta, p; \theta, r)$ to sum U, provided $p \geqq 0$, $r \geqq 0$. Both $s^{-p}t^{-r}C^{pr}_{\lambda\mu}(s, t)$ and $s^{-\rho}t^{-\tau}C^{\rho\tau}_{\eta\theta}(s, t)$ are bounded.

This is a special case of the following theorem.

THEOREM VII′. *If the series $\sum u_{kl}$ is summable $(R: \lambda, p; \mu, r)$ to sum U, and if $C^{pr}_{\lambda\mu}(s, t) = O(s^p t^r)$; if η is an L-function of λ, and θ an L-function of μ, such that*

$$\eta = O(\lambda^{\Delta_1}) \text{ and } \theta = O(\mu^{\Delta_2}),$$

where Δ_1 and Δ_2 are constants, then $\sum u_{kl}$ is summable $(R: \eta, p; \theta, r)$ to sum U, and $C^{\rho\tau}_{\eta\theta}(s, t) = O(s^p t^r)$, provided $p \geqq 0$, $r \geqq 0$.

V. NECESSARY CONDITIONS FOR SUMMABILITY

THEOREM VIII. *If $\sum_{k,l=1,\cdots,\infty} u_{kl}$ is summable $(R: \lambda, p; \mu, r)$ to sum U, then*

$$U_{kl} - U = o\left[\left(\frac{\lambda_{k+1}}{\lambda_{k+1} - \lambda_k}\right)^p \left(\frac{\mu_{l+1}}{\mu_{l+1} - \mu_l}\right)^r\right].$$

If, in addition, $C^{pr}_{\lambda\mu}(s, t) = O_s(t^r)$, then

$$U_{kl} - U = O_k\left[\left(\frac{\mu_{l+1}}{\mu_{l+1} - \mu_l}\right)^r\right];$$

if $C^{pr}_{\lambda\mu}(s, t) = O_t(s^p)$, then

$$U_{kl} - U = O_l\left[\left(\frac{\lambda_{k+1}}{\lambda_{k+1} - \lambda_k}\right)^p\right].$$

This theorem is a special case of the following more general theorem.

THEOREM IX. *If $\sum_{k,l=1,\cdots,\infty} u_{kl}$ is summable $(R: \lambda, p; \mu, r)$ to sum U, and if p' and r' are integers such that $0 \leqq p' \leqq p$, $0 \leqq r' \leqq r$, where $\lambda_k < s$ for $p' = 0$, $\mu_l < t$ for $r' = 0$, $\lambda_k \leqq s \leqq \lambda_{k+1}$, for $p' > 0$, and $\mu_l \leqq t \leqq \mu_{l+1}$ for $r' > 0$, then*

$$C^{p'r'}_{\lambda\mu}(s, t) - Us^{p'}t^{r'} = o\left[\frac{\lambda^p_{k+1}}{(\lambda_{k+1} - \lambda_k)^{p-p'}} \cdot \frac{\mu^r_{l+1}}{(\mu_{l+1} - \mu_l)^{r-r'}}\right].$$

If, in addition, $C^{pr}_{\lambda\mu}(s, t) = O_s(t^r)$, then

$$C^{p'r'}_{\lambda\mu}(s, t) - Us^{p'}t^{r'} = O_s\left[\frac{\mu^r_{l+1}}{(\mu_{l+1} - \mu_l)^{r-r'}}\right];$$

if $C^{pr}_{\lambda\mu}(s, t) = O_t(s^p)$, then

$$C^{p'r'}_{\lambda\mu}(s, t) - Us^{p'}t^{r'} = O_t\left[\frac{\lambda^p_{k+1}}{(\lambda_{k+1} - \lambda_k)^{p-p'}}\right].$$

* Cf. M., p. 521, Theorem II (i). The proof is incorrect when p and r (or either p or r) are integral, since under these conditions the formula for $C^{r,s}(X, \Omega)$ does not follow from Lemma 1.

Case 1. Assume p and r integral; $p \geqq 1, r \geqq 1$.

Assume $U=0$, and u_{kl} real.

Let $\Omega_x - \lambda_k = xh$, where $x=1, 2, \cdots, p$ and $h=(\lambda_{k+1}-\lambda_k)/p$, and let $\psi_y - \mu_l = yj$, where $y=1, 2, \cdots, r$ and $j=(\mu_{l+1}-\mu_l)/r$.

Let Ω and ψ denote any of these numbers.

By Lemma 3, applied twice, we have

$$(5.11)\quad pr\int_0^{\Omega}\int_0^{\psi} C_{\lambda\mu}^{p-1,r-1}(\sigma,\tau)\,d\tau d\sigma = C_{\lambda\mu}^{pr}(\Omega,\psi) = o(\Omega^p\psi^r) = o(\lambda_k^p{}_{+1}\mu_l^r{}_{+1}).$$

Using Lemma 7, and the fact that

$$C_{\lambda\mu}^{pr}(\sigma,\tau) = o(\sigma^p\tau^r) = o(\lambda_k^p{}_{+1}\mu_l^r{}_{+1}),$$

we have

$$\int_{\lambda_k}^{\Omega}\int_{\mu_l}^{\psi} C_{\lambda\mu}^{p-1,r-1}(\sigma,\tau)\,d\tau d\sigma = \int_0^{\Omega}\int_0^{\psi} - \int_0^{\lambda_k}\int_0^{\psi} - \int_0^{\Omega}\int_0^{\mu_l} + \int_0^{\lambda_k}\int_0^{\mu_l} = o(\nu_k^p{}_{+1}\mu_l^r{}_{+1}).$$

Writing

$$\int_{\lambda_k}^{\Omega}\int_{\mu_l}^{\psi} C_{\lambda\mu}^{p-1,r-1}(\sigma,\tau)\,d\tau d\sigma = \int_{\lambda_k}^{\Omega}\int_{\mu_l}^{\psi} C_{\lambda\mu}^{p-1,r-1}(\sigma,\tau)\,d(\psi-\tau)\,d(\Omega-\sigma)$$

and integrating $(r-1)$ times by parts with respect to τ, and $(p-1)$ times with respect to σ, we have

$$\begin{aligned}(5.2)\qquad &(\psi-\mu_l)\left[(\Omega-\lambda_k)\,C_{\lambda\mu}^{p-1,r-1}(\lambda_k,\mu_l) + \frac{p-1}{2}(\Omega-\lambda_k)^2 C_{\lambda\mu}^{p-2,r-1}(\lambda_k,\mu_l)\right.\\ &\left. + \cdots + \frac{1}{p}(\Omega-\lambda_k)^p C_{\lambda\mu}^{0,r-1}(\lambda_k+0,\mu_l)\right]\\ &+ \frac{r-1}{2}(\psi-\mu_l)^2\left[(\Omega-\lambda_k)\,C_{\lambda\mu}^{p-1,r-2}(\lambda_k,\mu_l)\right.\\ &\left. + \cdots + \frac{1}{p}(\Omega-\lambda_k)^p C_{\lambda\mu}^{0,r-2}(\lambda_k+0,\mu_l)\right] + \cdots\\ &+ \frac{1}{r}(\psi-\mu_l{}^r)\left[(\Omega-\lambda_k)C_{\lambda\mu}^{p-1,0}(\lambda_k,\mu_l+0)\right.\\ &\left. + \cdots + \frac{1}{p}(\Omega-\lambda_k)^p C_{\lambda\mu}^{00}(\lambda_k+0,\mu_l+0)\right].\end{aligned}$$

For the p different values of x, and the r different values of y, we get pr different linear combinations of the form (5.2). The proof of the first part of the theorem may be completed as in M., p. 523.

The proof of the second part of the theorem is entirely similar if o is replaced by O_s or by O_t.

Case 2. Assume p and r not integral. The proof is that of M., p. 522.

Case 3. Assume p integral, and r not integral.

We define Ω_x as in Case 1, and ψ_y as in Case 2; we apply Lemma 3 twice, and then Lemma 7, and integrate the result by parts, $[r]$ times with respect to τ and $(p-1)$ times with respect to σ. The rest of the proof is similar to that for Case 1.

By a similar proof, we can show that the theorem is true for p not integral, and r integral.

By means of Theorem VIII, we can prove the following theorems.

THEOREM X. *If* $\lambda_m = O(\lambda_m - \lambda_{m-1})$, *then no double series can be summable* $(R: \lambda, p; \mu, 0)$ *unless it is convergent.*

THEOREM XI. *If* $\mu_n = O(\mu_n - \mu_{n-1})$, *then no double series can be summable* $(R: \lambda, 0; \mu, r)$ *unless it is convergent.*

THEOREM XII. *If* $\lambda_m = O(\lambda_m - \lambda_{m-1})$ *and* $\mu_n = O(\mu_n - \mu_{n-1})$, *then no double series can be summable* $(R: \lambda, p; \mu, r)$ *unless it is convergent.*

VI. MULTIPLICATION

We consider the double series $\sum_{m,n=1,\cdots,\infty} a_{mn}$ and $\sum_{m,n=1,\cdots,\infty} b_{mn}$, and form the Dirichlet double series

$$\sum_{1,1}^{\infty,\infty} a_{mn}\, e^{-\lambda_{m'}s-\mu_{n'}t} \quad \text{and} \quad \sum b_{mn}\, e^{-\lambda_{m''}s-\mu_{n''}t}.$$

Let λ_i be the ascending sequence formed by all the values of $\lambda_p'+\lambda_q''$, and μ_j the ascending sequence formed by all the values of $\mu_{p'}'+\mu_{q'}''$. If $\lambda_{p_i}'+\lambda_{q_i}''=\lambda_{p_s}'+\lambda_{q_s}''$, so that two values of λ are the same, the order of these two values is indifferent. The same thing is true for two equal values of μ.

The series $\sum_{i,j=1,\cdots,\infty} c_{ij}$, where

$$c_{ij} = \sum_{\lambda_n'+\lambda_k''=\lambda_i} \sum_{\mu_n'+\mu_l''=\mu_j} a_{mn}b_{kl},$$

is analogous to the Dirichlet product for simple series. We shall call it the Dirichlet product of the double series $\sum a_{mn}$, $\sum b_{mn}$, of type $(\lambda', \mu'; \lambda'', \mu'')$.

We shall call the double series $\sum_{m,n=1,\cdots,\infty} c_{mn}$, where

$$\begin{aligned} c_{mn} = \quad & a_{11}b_{mn} \quad + \cdots + a_{1n}b_{m1} \\ & + a_{21}b_{m-1,n} + \cdots + a_{2n}b_{m-1,1} + \cdots \\ & + a_{m1}b_{1n} \quad + \cdots + a_{mn}b_{11}, \end{aligned}$$

the Cauchy product of the double series $\sum a_{mn}$ and $\sum b_{mn}$. It is analogous to the Cauchy product for simple series.

The analogues of the theorems of Mertens, Abel and Cauchy for the Cauchy product of simple series have been proved for the Dirichlet product of simple series.* We shall prove corresponding theorems for the Dirichlet product of double series. Since the Cauchy product is a special case of the Dirichlet product, it follows that these theorems must be true also for the Cauchy product of double series.

THEOREM XIII. *If $\sum a_{mn}$ is absolutely convergent to sum A, and $\sum b_{mn}$ convergent to sum B and bounded, then $\sum c_{ij}$, the Dirichlet product of the type $(\lambda', \mu'; \lambda'', \mu'')$, is convergent to sum C and bounded, and $C = AB$.*

We assume that $B = 0$.

Let $C_{\mu\nu}$ be a partial sum of the product series, and let a_{mn} be the a of highest rank in it. Then

$$c_{\mu\nu} = \sum_{i=1,j=1}^{m,n} a_{ij}B_{kl}, \quad \text{where } B_{kl} = \sum_{i=1,j=1}^{k,l} b_{ij}$$

and where k is a function of i and m, and l a function of j and n.

Suppose $C_{\mu\nu}$ contains a term $a_{\gamma\delta}b_{\gamma\delta}$. Then it contains all the terms $a_{pq}b_{rs}$ $(p \leqq \gamma, r \leqq \gamma, q \leqq \delta, s \leqq \delta)$. Therefore

$$\begin{aligned} & m \geqq \gamma, \qquad n \geqq \delta; \\ & k \geqq \gamma, \quad \text{for } i = 1, 2, \cdots, \gamma, \\ & l \geqq \delta, \quad \text{for } j = 1, 2, \cdots, \delta. \end{aligned}$$

Since $\sum b_{mn} \to 0$, and since $\sum a_{mn}$ is absolutely convergent, it is possible to choose γ and δ so that

$$|B_{kl}| < \epsilon, \quad k \geqq \gamma, l \geqq \delta$$

and

$$\sum_{1,1}^{\infty,\infty} |a_{ij}| - \sum_{1,1}^{\gamma,\delta} |a_{ij}| < \epsilon.$$

---

* H. R., pp. 63, 64.

By hypothesis, B_{kl} is bounded. Let $\overline{B}$ be an upper bound of $|B_{kl}|$. Then

$$|C_{\mu\nu}| \leqq \overline{B}\left[\sum_{1,1}^{m,n}|a_{ij}| - \sum_{1,1}^{\gamma,\delta}|a_{ij}|\right] + \epsilon\sum_{1,1}^{\gamma,\delta}|a_{ij}|$$
$$< \epsilon(\overline{B} + A).$$

Therefore $C_{\mu\nu}$ approaches zero as γ and δ approach ∞, that is, as m and n approach ∞.

If $B \neq 0$, we form a new series, $\sum b'_{mn}$, such that

$$b'_{11} = b_{11} - B$$

and

$$b'_{mn} = b_{mn}, \quad \text{for } m \neq 1 \text{ and } n \neq 1.$$

Then the Dirichlet product of $\sum a_{mn}$ and $\sum b'_{mn} \to 0$; therefore the Dirichlet product of $\sum a_{mn}$ and $\sum b_{mn} \to AB$.

To prove that the product series is bounded, we prove first that since $\sum a_{mn}$ is absolutely convergent, $\sum |a_{mn}|$ is bounded. For there exists an M, such that when $m > M$, $n > M$,

$$\sum_{1,1}^{m,n}|a_{ij}| < A + \epsilon.$$

But for any p and q we can find an m and an n, such that $p < m$, $q < n$, $m > M$, $n > M$. Therefore

$$\sum_{1,1}^{p,q}|a_{ij}| < \sum_{1,1}^{m,n}|a_{ij}| < A + \epsilon, \text{ for all } p \text{ and } q,$$

and $\sum |a_{mn}|$ is bounded.

Let A be an upper bound of $\sum |a_{mn}|$. We have

$$C_{\mu\nu} = \sum_{i=1,j=1}^{m,n} a_{ij}B_{kl}.$$

Therefore

$$\sum_{i=1,j=1}^{m,n} |a_{ij}||B_k| \leqq \overline{B}\,\overline{A}$$

THEOREM XIV. *If $\sum a_{mn}$, $\sum b_{mn}$ and $\sum c_{mn}$ are all convergent and bounded, where $\sum c_{mn}$ is the Dirichlet product, then $AB = C$.*

By means of Theorem IV, we deduce this theorem from the following more general theorem.

THEOREM XV. *If $\sum a_{mn}$ is summable $(R: \lambda', p'; \mu', r')$ to sum A, and if $\sum b_{mn}$ is summable $(R: \lambda'', p''; \mu'', r'')$ to sum B, where p', r', p'' and $r'' > -1$, and if $\sigma^{p'}\tau^{r'}A^{p'r'}_{\lambda'\mu'}(s,t)$ and $\sigma^{p''}\tau^{r''}B^{p''r''}_{\lambda''\mu''}(s,t)$ are bounded, then the Dirichlet product series, $\sum c_{mn}$, is summable $(R: \lambda, p; \mu, r)$ to sum C, where $p = p' + p'' + 1, r = r' + r'' + 1$, and $AB = C$, and $C^{pr}_{\lambda\mu}(s,t)$ is bounded.*

We have

$$A^{p'r'}_{\lambda'\mu'}(s,t) = \sum_{\lambda'_m < s} \sum_{\mu'_n < t} (s - \lambda'_m)^{p'}(t - \mu_n)^{r'} a_{mn},$$

$$B^{p''r''}_{\lambda''\mu''}(s,t) = \sum_{\lambda_k'' < s} \sum_{\mu_l'' < t} (s - \lambda_k'')^{p''}(t - \mu_l'')^{r''} b_{kl},$$

and

$$C^{pr}_{\lambda\mu}(s,t) = \sum_{\lambda_i < s} \sum_{\mu_j < t} (s - \lambda_i)^{p}(t - \mu_j)^{r} c_{ij}.$$

We shall prove that

$$(6.1) \qquad C^{pr}_{\lambda\mu}(s,t) = \frac{\Gamma(p+1)}{\Gamma(p'+1)\Gamma(p''+1)} \cdot \frac{\Gamma(r+1)}{\Gamma(r'+1)\Gamma(r''+1)} \cdot \int_0^s \int_0^t A^{p'r'}_{\lambda'\mu'}(\sigma, \tau) B^{p''r''}_{\lambda''\mu''}(s-\sigma, t-\tau) d\tau d\sigma.$$

For consider the term $a_{mn}b_{kl}$. It occurs in $C^{pr}_{\lambda\mu}(s, t)$ if $\lambda_m' + \lambda_k'' < s$ and $\mu_n' + \mu_l'' < t$, with the coefficient

$$(s - \lambda_i)^{p}(t - \mu_j)^{r} = (s - \lambda'_m - \lambda_k'')^{p}(t - \mu_n' - \mu_l'')^{r}.$$

The term a_{mn} occurs in $A^{p'r'}_{\lambda'\mu'}(\sigma, \tau)$ if $\lambda_m' < \sigma$, and $\mu_n' < \tau$, with the coefficient $(\sigma - \lambda_m')^{p'}(\tau - \mu_n')^{r'}$; b_{kl} occurs in $B^{p''r''}_{\lambda''\mu''}(s-\sigma, t-\tau)$ if $\lambda_k'' < s - \sigma$, and $\mu_l'' < t - \tau$, with the coefficient $(s - \sigma - \lambda_k'')^{p''}\ (t - \tau - \mu_l'')^{r''}$. Therefore $a_{mn}b_{kl}$ occurs in the right hand member of (6.1) if $\lambda_m' + \lambda_k'' < s$ and $\mu_n' + \mu_l'' < t$, and it scoefficient is

$$\frac{\Gamma(p+1)}{\Gamma(p'+1)\Gamma(p''+1)} \cdot \frac{\Gamma(r+1)}{\Gamma(r'+1)\Gamma(r''+1)} \cdot \int_{\lambda'_m}^{s-\lambda_k''} \int_{\mu'_n}^{t-\mu_l''} (\sigma - \lambda'_m)^{p'}(\tau - \mu_n)^{r'}(s - \sigma - \lambda_k'')^{p''}(t - \tau - \mu_l'')^{r''} d\tau d\sigma$$

$$= \frac{\Gamma(p+1)}{\Gamma(p'+1)\Gamma(p''+1)} \cdot \frac{\Gamma(r+1)}{\Gamma(r'+1)\Gamma(r''+1)} \int_0^x \int_0^y u^{p'} v^{r'} (x-u)^{p''} (y-v)^{r''} dv du$$

$$= x^{p'+p''+1}y^{r'+r''+1} = (s - \lambda_m' - \lambda_k'')^p(t - \mu_n' - \mu_l'')^r.$$

Therefore (6.1) is established.

But $A^{p'r'}_{\lambda'\mu'}(\sigma, \tau) \sim A\sigma^{p'}\tau^{r'}$ and $B^{p''r''}_{\lambda''\mu''}(\sigma, \tau) \sim B\sigma^{p''}\tau^{r''}$, and $A^{p'r'}_{\lambda'\mu'}(\sigma, \tau)/(\sigma^{p'}\tau^{r'})$ and $B^{p''r''}_{\lambda''\mu''}(\sigma, \tau)/(\sigma^{p''}\tau^{r''})$ are bounded.

Therefore by Lemma 2, $C^{pr}_{\lambda\mu}(s, t) \sim ABs^pt^r$, and the transformed sequence is bounded, which proves the theorem.

THEOREM XVI. *If $\sum a_{mn}$ and $\sum b_{mn}$ are absolutely convergent with sums A and B, then the Dirichlet product series of type $(\lambda', \mu'; \lambda'', \mu'')$ is absolutely convergent with sum AB.*

Since $\sum a_{mn}$ and $\sum b_{mn}$ are absolutely convergent, the series $\sum|a_{mn}|$ and $\sum|b_{mn}|$ are convergent and bounded. Therefore, by Theorem XIII, their product series is convergent and bounded.

Suppose

$$c_{ij} = \sum_{\lambda_m'+\lambda_k''=\lambda_i} \sum_{\mu_n'+\mu_l''=\mu_j} a_{mn}b_{kl}$$

and

$$d_{ij} = \sum_{\lambda_m'+\lambda_k''=\lambda_i} \sum_{\mu_n'+\mu_l''=\mu_j} a_{mn}b_{kl}.$$

Then $|d_{ij}| \leqq c_{ij}$, and by the comparison test for double series, $\sum|d_{ij}|$ is convergent. Hence $\sum d_{ij}$ is absolutely convergent and by Theorem XIII its sum is AB.

CORNELL UNIVERSITY,
ITHACA, N. Y.

THE INVERSE PROBLEM OF THE CALCULUS OF VARIATIONS IN HIGHER SPACE*

BY
DAVID R. DAVIS

INTRODUCTION

The chief problem of this paper is to determine the integrand f for an integral I of the form

$$I = \int_{x_1}^{x_2} f(x, y, z, y', z')dx$$

whose Euler equations are

$$H(x, y, z, y', z', y'', z'') = 0, \qquad K(x, y, z, y', z', y'', z'') = 0.$$

The problem is impossible unless H and K have special properties. In determining these properties it is found to be necessary and sufficient that the equations of variation,

$$H_y u + H_z v + H_{y'} u' + H_{z'} v' + H_{y''} u'' + H_{z''} v'' = 0,$$
$$K_y u + K_z v + K_{y'} u' + K_{z'} v' + K_{y''} u'' + K_{z''} v'' = 0,$$

of $H = K = 0$ be a self-adjoint system.

In §I systems of adjoint differential equations of the second order are discussed. The unique existence of an adjoint system is proved and necessary and sufficient conditions for a given system of the second order to be self-adjoint are established.

Characteristic properties of the differential equations of the calculus of variations are derived in §II. It is here shown that if two equations $H = K = 0$ are to be the Euler equations for an integral I of the form given above they must have their equations of variation self-adjoint.

The sufficiency of this property is established in §III. Necessary and sufficient conditions for the equations of variation of the system $H = K = 0$ to be self-adjoint are given. Finally, the most general form of the integrand f for the integral I whose Euler equations are the equations $H = K = 0$ is determined.

---

* Presented to the Society, San Francisco Section, June 18, 1927, under the titles *Characteristic properties of differential equations of the calculus of variations*, *Integrals whose extremals have given differential equations*, and *A method of determining an integral I whose extremals are a given family of curves*; received by the editors June 1, 1927.

The problem of finding an integral whose extremals are a given family of curves is treated in §IV. Necessary and sufficient conditions for two given equations of the form

$$y'' = F(x,y,z,y',z'), \qquad z'' = G(x,y,z,y',z')$$

to have solutions $y=y(x)$, $z=z(x)$ which represent the extremals for an integral I of the form above are derived. Also the possibility of determining linear combinations of two given equations $y''=F$, $z''=G$, which shall have equations of variation that are self-adjoint, is discussed.

In the last three sections special examples are treated. In §V is found the most general type of integral whose extremals are straight lines. In §VI is treated the problem of finding integrals whose extremals are semi-circles orthogonal to the xy-plane and finally, in §VII, that of determining integrals whose extremals are catenaries orthogonal to the xy-plane.*

I. Self-adjoint systems of linear differential expressions of the second order

The properties which characterize a self-adjoint system of differential expressions of the second order are needed in the discussion of the following sections. For convenience of reference they are developed here.†

Consider the system of differential expressions

$$(1) \qquad J_i(u) = A_{ik}(x)u_k + B_{ik}(x)u_k' + C_{ik}(x)u_k'' \qquad (i = 1, \cdots, n),$$

in the n variables u_k $(k=1, \cdots, n)$. In this and following expressions the notation of tensor analysis is used, that is, whenever two subscripts are alike in a term, say of the form $A_{ik}u_k$, then the expression represents a sum with respect to the repeated index. The use of accents or primes will always indicate the total derivative of the function with respect to the independent variable.

A system $K_i(u)$ of the same form as (1) and satisfying with $J_i(u)$ an equation

$$(2) \qquad u_iJ_i(v) - v_iK_i(u) = \frac{dM}{dx},$$

---

* The writer wishes to take this opportunity to thank Dr. Bliss of Chicago for his advice and valuable suggestions during the preparation of this paper.

† See J. Kürschak, *Die Existenzbedingungen des verallgemeinerten kinetischen Potentials*, Mathematische Annalen, vol. 62 (1906), p. 148. Also G. Frobenius, *Über adjungierte lineare Differentialausdrücke*, Journal für Mathematik, vol. 85 (1878), p. 207.

where M is a function of x, u_i, v_k and their derivatives, is called an adjoint system of $J_i(u)$.

It is easily proved that every system $J_i(u)$ has one and only one adjoint system $K_i(u)$. For from (1), we have

$$\begin{aligned} u_iJ_i(v) &= u_iA_{ik}v_k + u_iB_{ik}v_k' + u_iC_{ik}v_k'' \\ &= u_iA_{ik}v_k + \frac{d}{dx}(u_iB_{ik}v_k) + \frac{d}{dx}[(u_iC_{ik})v_k' - (u_iC_{ik})'v_k] \\ &\quad - (u_iB_{ik})'v_k + (u_iC_{ik})''v_k, \\ &= [u_iA_{ik} - (u_iB_{ik})' + (u_iC_{ik})'']v_k \\ &\quad + \frac{d}{dx}[u_iB_{ik}v_k + (u_iC_{ik})v_k' - (u_iC_{ik})'v_k]. \end{aligned}$$

Hence, there exist a system $K_i(u)$ and a function M which satisfy equation (2), namely,

$$\begin{aligned} K_i(u) &= u_iA_{ik} - (u_iB_{ik})' + (u_iC_{ik})'', \\ M &= u_iB_{ik}v_k + (u_iC_{ik})v_k' - (u_iC_{ik})'v_k. \end{aligned} \tag{3}$$

Suppose there were two such systems; then we would have the two equations

$$u_iJ_i(v) - v_iK_i(u) = \frac{dM}{dx},$$

$$u_iJ_i(v) - v_iH_i(u) = \frac{dL}{dx},$$

which give

$$v_i[H_i(u) - K_i(u)] = \frac{d}{dx}(M - L).$$

If we represent the coefficients of the v_i for a fixed set $u_i(x)$ by the notations $\phi_i(x)$, then the integral

$$\int_{x_1}^{x_2} v_i\phi_i(x)dx = M - L\Big|_{x_1}^{x_2}$$

is clearly independent of the path in the xv-plane. This can be so, however, if and only if the functions $\phi_i(x)$ are identically zero, which implies that for every set of functions $u_i(x)$

$$H_i(u) \equiv K_i(u).$$

Thus the adjoint system $K_i(u)$ as defined by (2) is unique.

The $J_i(u)$ are said to be self-adjoint if the systems $J_i(u)$ and $K_i(u)$ are identical. Imposing this condition, we have by (1) and (3)

$$A_{ik}u_k + B_{ik}u_k' + C_{ik}u_k'' \equiv u_kA_{ki} - (u_kB_{ki})' + (u_kC_{ki})''$$
$$\equiv u_k(A_{ki} - B_{ki}' + C_{ki}'') + u_k'(2C_{ki}' - B_{ki}) + u_k''C_{ki}.$$

Regarding these equations as identities in the u_k and their derivatives and equating the corresponding coefficients, we arrive at the following theorem:

THEOREM I. *Necessary and sufficient conditions that the system of differential expressions*

$$J_i(u) = A_{ik}(x)u_k + B_{ik}(x)u_k' + C_{ik}(x)u_k'' \qquad (i = 1, \cdots, n)$$

shall be self-adjoint are

$$C_{ik} \equiv C_{ki}, \qquad B_{ik} + B_{ki} \equiv 2C_{ik}', \qquad A_{ik} \equiv A_{ki} - B_{ki}' + C_{ki}''. \tag{4}$$

II. CHARACTERISTIC PROPERTIES OF DIFFERENTIAL EQUATIONS OF THE CALCULUS OF VARIATIONS*

In the problem of the calculus of variations in a space of $n+1$ dimensions one seeks to minimize the integral

$$I = \int_{x_1}^{x_2} f(x, y_1, \cdots, y_n, y_1', \cdots, y_n')dx$$

in a class of arcs $y_k = y_k(x)$, $x_1 \leqq x \leqq x_2$, joining two given points. It is well known that if a minimum is to occur it is necessary that the first variation of the integral I must vanish and that the second variation must be always positive or zero. The vanishing of the first variation implies that the Euler-Lagrange equations

$$\frac{d}{dx} f_{y'} - f_y = 0 \tag{5}$$

must be satisfied along a minimizing arc. The subscripts y and y' are used to represent y_i and y_i' respectively. If the equations

$$y_i = y_i(x, b) \qquad (i = 1, \cdots, n) \tag{6}$$

define a one-parameter family of solutions of the Euler-Lagrange equations, then the functions

$$u_i = y_{ib}(x, b)$$

are called the variations of the family along the arc (6). Along this arc they must satisfy the so-called equations of variation

$$J_i(u) = \frac{d}{dx}(f_{i'k}u_k + f_{i'k'}u_k') - (f_{ik}u_k + f_{ik'}u_k') = 0$$

* See J. Hadamard, *Leçons sur le Calcul des Variations*, p. 151.

which are found by differentiating equations (5) with respect to the parameter b which occurs in the arguments y_i and y_i' of the function f. The subscripts of f denote differentiation with respect to the corresponding y's.

Let the function ω be defined by the expression

$$2\omega = f_{ik}u_iu_k + 2f_{ik'}u_iu_k' + f_{i'k'}u_i'u_k' ;$$

then it is readily verified that

$$J_i(u) = \frac{d}{dx}\omega_{u'} - \omega_u \qquad (u' = u_i',\ u = u_i). \tag{7}$$

It is now easy to show that the system $J_i(u)$ is self-adjoint.* According to the definition of §I we must have

$$u_iJ_i(v) - v_iJ_i(u) = \frac{dM}{dx}.$$

With the aid of (7) we readily find that

$$u_iJ_i(v) - v_iJ_i(u) = \frac{d}{dx}(u_i\omega_{v'} - v_i\omega_{u'}),$$

in which the subscripts v' and u' represent v_i' and u_i' respectively. Therefore a set of differential equations which are the differential equations of the solutions of a problem of the calculus of variations must have the property indicated in the following theorem:

THEOREM II. *If a given system of differential equations*

$$H_j(x, y_1, \cdots, y_n, y_1', \cdots, y_n', y_1'', \cdots, y_n'') = 0 \quad (j = 1, \cdots, n)$$

is to be the system of differential equations of the solutions of the problem of minimizing the integral I, of the type indicated above, then their equations of variation

$$H_{ji}u_i + H_{ji'}u_i' + H_{ji''}u_i'' = 0$$

must be a self-adjoint system along every curve $y_i = y_i(x)$.

III. INTEGRALS WHOSE EXTREMALS HAVE GIVEN DIFFERENTIAL EQUATIONS

Consider the equations

$$H(x,y,z,y',z',y'',z'') = 0, \qquad K(x,y,z,y',z',y'',z'') = 0, \tag{8}$$

whose solutions have the form

$$y = y(x), \quad z = z(x)$$

* It is of course understood that the argument here is concerning the Euler equations as they stand and not an equivalent system.

with the derivatives y', z', y'', z''. It will be shown, upon the assumption that the equations of variation of $H=K=0$ form a self-adjoint system, that a function $f(x, y, z, y', z')$ can be determined such that the equations (8) are the differential equations of the solutions of the problem of minimizing the integral

$$I = \int_{x_1}^{x_2} f(x,y,z,y',z')dx. \tag{9}$$

The equations of variation of the given system (8) are

$$\begin{aligned} &H_y u + H_z v + H_{y'}u' + H_{z'}v' + H_{y''}u'' + H_{z''}v'' = 0, \\ &K_y u + K_z v + K_{y'}u' + K_{z'}v' + K_{y''}u'' + K_{z''}v'' = 0. \end{aligned} \tag{10}$$

If these equations are a self-adjoint system, then by Theorem I, we have the following relations:

$$\begin{aligned} &H_{z''} = K_{y''}, \quad H_{y'} = H'_{y''}, \quad K_{z'} = K'_{z''}, \\ &H_{z'} + K_{y'} = 2H'_{z''} = 2K'_{y''}, \\ &K_y = H_z - H'_{z'} + H''_{z''}, \quad H_z = K_y - K'_{y'} + K''_{y''}, \end{aligned} \tag{11}$$

which must be identities in $x, y, z, y', z', y'', z'', y''', z'''$. The last of equations (11) is a consequence of the preceding relations. For we may solve the fifth equation for H_z and make use of relations (1) and (4) of (11). Then

$$\begin{aligned} H_z &= K_y + H'_{z'} - H''_{z''}, \\ &= K_y + 2K''_{y''} - K'_{y'} - K''_{y''}, \end{aligned}$$

which gives the last of equations (11).

In the second and third of relations (11) terms in y''' and z''' do not occur in the first members. It is thus evident upon expanding the second members that the derivatives $H_{y''y''}$; $H_{y''z''}$; $K_{z''y''}$; $K_{z''z''}$ all vanish identically, and hence, that the given equations (8) are linear in y'' and z''. Also the first equation of (11) asserts that the coefficient of z'' in (8_1) is equal to the coefficient of y'' in (8_2). Therefore the given equations may be expressed in the form

$$\begin{aligned} H &= M + Py'' + Qz'' = 0, \\ K &= N + Qy'' + Rz'' = 0, \end{aligned} \tag{12}$$

where M, N, P, Q, R are functions of x, y, z, y', z'.

In view of the first of relations (11) and upon subtracting the fifth from the last, one obtains

$$2(H_z - K_y) = H'_{z'} - K'_{y'} = \frac{d}{dx}(H_{z'} - K_{y'}).$$

Expanding the second member of this equation and noting that terms in y''' and z''' do not occur on the left, we have

$$H_{z'y''} - K_{y'y''} = 0, \qquad H_{z'z''} - K_{y'z''} = 0.$$

Hence, by (11_1) each of the functions H and K must satisfy the conditions

$$H_{y'z''} = H_{y''z'}, \qquad K_{y'z''} = K_{y''z'},$$

which in terms of the notation in equations (12) become

$$P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'}. \tag{13}$$

These properties of the given functions H and K are the necessary conditions of integrability which are later required in the determination of the integrand f.

With the aid of (12) and (13) the second, third and fourth of the self-adjoint relations (11) may be written in the form

$$\begin{aligned} M_{y'} &= P_z + P_y y' + P_z z', \\ N_{z'} &= R_z + R_y y' + R_z z', \\ M_{z'} + N_{y'} &= 2(Q_z + Q_y y' + Q_z z'). \end{aligned} \tag{14}$$

Since the last two of relations (11) are equivalent, the use of either together with equations (12) will give

$$M_z - N_y = (Q_y - P_z)y'' + (R_y - Q_z)z'' + \tfrac{1}{2}(M_{z'} - N_{y'})'.$$

If this equation is to be an identity in the variables involved, we must have

$$\begin{aligned} Q_y - P_z &= -\tfrac{1}{2}(M_{z'y'} - N_{y'y'}), \\ R_y - Q_z &= -\tfrac{1}{2}(M_{z'z'} - N_{y'z'}), \\ M_z - N_y &= \tfrac{1}{2}\left[(M_{z'} - N_{y'})_z + (M_{z'} - N_{y'})_y y' + (M_{z'} - N_{y'})_z z'\right]. \end{aligned} \tag{15}$$

The first two of these relations may be obtained without difficulty from (14) and (13). But the third involves an independent relation which must be satisfied by the functions M and N. We can now state the following theorem:

THEOREM III. *If a system of equations*

$$H(x,y,z,y',z',y'',z'') = 0, \qquad K(x,y,z,y',z',y'',z'') = 0$$

is to have equations of variation

$$\begin{aligned} H_y u + H_z v + H_{y'} u' + H_{z'} v' + H_{y''} u'' + H_{z''} v'' &= 0, \\ K_y u + K_z v + K_{y'} u' + K_{z'} v' + K_{y''} u'' + K_{z''} v'' &= 0, \end{aligned}$$

which are self-adjoint along every curve $y=y(x)$, $z=z(x)$, *then it must have the form*

$$H = M(x,y,z,y',z') + P(x,y,z,y',z')y'' + Q(x,y,z,y',z')z'',$$
$$K = N(x,y,z,y',z') + Q(x,y,z,y',z')y'' + R(x,y,z,y',z')z'',$$

where M, N, P, Q, R *satisfy the following relations*:

$$\begin{aligned} &P_{z'} = Q_{y'}, \quad Q_{z'} = R_{y'}, \\ &M_{y'} = P_z + P_y y' + P_z z', \quad N_{z'} = R_z + R_y y' + R_z z', \\ &M_{z'} + N_{y'} = 2(Q_z + Q_y y' + Q_z z'), \\ &M_z - N_y = \tfrac{1}{2}\left[(M_{z'} - N_{y'})_z + (M_{z'} - N_{y'})_y y' + (M_{z'} - N_{y'})_z z'\right], \end{aligned} \tag{16}$$

identically in x, y, z, y', z'.

If a system of equations $H=K=0$ has the form indicated in the last theorem with coefficients satisfying (16), then it is possible to determine the integrand of the integral I in (9) to an arbitrary function of x, y, z so that $H=K=0$ are the equations of its minimizing curves. Let us first determine a function $g(x, y, z, y', z')$ satisfying the relations

$$g_{y'y'} = P, \qquad g_{y'z'} = Q, \qquad g_{z'z'} = R. \tag{17}$$

It is possible to do this since the conditions for integrability of the system (17) are the first two of the relations (16) for self-adjointness. The function g will then be determined by the integral

$$g = \int_{y_0'z_0'}^{y'z'} M\,dy' + L\,dz',$$

where the functions M and L represent the integrals

$$M = \int_{y_0'z_0'}^{y'z'} P\,dy' + Q\,dz', \quad L = \int_{y_0'z_0'}^{y'z'} Q\,dy' + R\,dz'.$$

The first two of relations (16) also insure that the above line integrals are independent of the path of integration. It is evident that if g is a particular solution of the system (17) then the most general one is given by the formula

$$f = g + a + by' + cz' \tag{18}$$

where a, b, c are constants with respect to the integrations for y' and z' and therefore functions of x, y, z.

If the given equations (8) are identical with the Euler-Lagrange equations for the integral (9), then the equations

$$(19)\qquad \frac{d}{dx}f_{y'} - f_y = M + Py'' + Qz'', \quad \frac{d}{dx}f_{z'} - f_z = N + Qy'' + Rz''$$

must be identities in $x, y, z, y', z', y'', z''$.

The value of f given in (18) must now satisfy equations (19). Inserting this value of f, we get

$$\frac{d}{dx}(g_{y'} + b) - (g_y + a_y + b_y y' + c_y z') = M + Py'' + Qz'',$$

$$\frac{d}{dx}(g_{z'} + c) - (g_z + a_z + b_z y' + c_z z') = N + Qy'' + Rz'',$$

which reduce to

$$(20)\qquad \begin{aligned}(b_x - a_y) + (b_z - c_y)z' &= M + g_y - g_{y'x} - g_{y'y}y' - g_{y'z}z',\\ (c_x - a_z) + (c_y - b_z)y' &= N + g_z - g_{z'x} - g_{z'y}y' - g_{z'z}z'.\end{aligned}$$

If these equations are to be identities in x, y, z, y', z' then the second members must be independent of y' and z' respectively, while the expression on the right in the first equation must be linear in z' and that of the second equation linear in y'. If the second member of the first equation is differentiated once with respect to y', the result by the third of relations (16) is identically zero. The second derivative of this same expression with respect to z' is readily shown to be identically zero by means of the fifth of (16) and the first two of (15), which are direct consequences of (16_3) and (16_4). One proceeds similarly in the case of the second equation. We thus have the required condition that a, b and c are functions of x, y, z alone.

From (20) one obtains the relations

$$(21)\qquad \begin{aligned}c_y - b_z &= g_{y'z} - g_{yz'} - \tfrac{1}{2}(M_{z'} - N_{y'}),\\ a_z - c_x &= -N - g_z + g_{z'x} + g_{y'z}y' + g_{z'z}z' - \tfrac{1}{2}(M_{z'} - N_{y'})y',\\ b_x - a_y &= M + g_y - g_{y'x} - g_{y'y}y' - g_{z'y}z' - \tfrac{1}{2}(M_{z'} - N_{y'})z'.\end{aligned}$$

The first one of these equations is found by differentiating the first or second equation of (20) with respect to y' or z' respectively and employing (16_5). The substitution of the value so found for $c_y - b_z$ in equations (20) gives the second and third of (21). The system (21) has solutions for a, b, c if and only if the second members, which from their derivation are independent of y' and z', satisfy the equation

$$(22)\qquad \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial r}{\partial z} \equiv 0$$

identically in x, y, z; where p, q, r represent the second members respectively. Performing the differentiation as indicated by (22) one finds that the terms in the derivatives of g cancel in pairs and that the resulting expression is identically zero by the last of relations (16).

Let a, b, c be a particular set of solutions of (21); then the most general solution will be

$$a + \alpha, \quad b + \beta, \quad c + \gamma,$$

where α, β, γ must satisfy the equations

$$\gamma_y - \beta_z = 0, \quad \alpha_z - \gamma_x = 0, \quad \beta_x - \alpha_y = 0$$

which are necessary and sufficient conditions for

$$\alpha + \beta y' + \gamma z'$$

to be the total derivative of a function $t(x, y, z)$.

The function f of (18) can now be expressed in the form

$$f = g + a + by' + cz' + \frac{dt}{dx}$$

where a, b, c are a particular set of solutions of (21), and t is an arbitrary function of (x, y, z). Hence, we have

THEOREM IV. *If two equations of the form*

$$H(x,y,z,y',z',y'',z'') = 0, \qquad K(x,y,z,y',z',y'',z'') = 0$$

have equations of variation

$$H_y u + H_z v + H_{y'} u' + H_{z'} v' + H_{y''} u'' + H_{z''} v'' = 0,$$
$$K_y u + K_z v + K_{y'} u' + K_{z'} v' + K_{y''} u'' + K_{z''} v'' = 0,$$

which form a self-adjoint system along every curve $y = y(x)$, $z = z(x)$, then there is always an integral of the form

$$I = \int_{x_1}^{x_2} f(x,y,z,y',z')dx$$

having the equations $H = K = 0$ as its Euler equations. The most general such integral has an integrand

$$f = g(x,y,z,y',z') + a(x,y,z) + b(x,y,z)y' + c(x,y,z)z' + \frac{d}{dx}t(x,y,z)$$

where g is a particular solution of (17); *a, b, c are a particular solution of* (21); *and t is an arbitrary function of x, y, z.*

IV. A METHOD OF DETERMINING AN INTEGRAL I WHOSE EXTREMALS ARE A GIVEN FAMILY OF CURVES

Consider the four-parameter family of arcs

$$y = y(x,a,b,c,d), \qquad z = z(x,a,b,c,d). \tag{23}$$

Differentiation with respect to x gives

$$y' = y_x(x,a,b,c,d), \qquad z' = z_x(x,a,b,c,d),$$
$$y'' = y_{xx}(x,a,b,c,d), \qquad z'' = z_{xx}(x,a,b,c,d).$$

If a, b, c, d are eliminated from the above equations one obtains two equations of the form

$$y'' = F(x,y,z,y',z'), \qquad z'' = G(x,y,z,y',z'). \tag{24}$$

The general solutions of these equations are (23) where a, b, c, d are constants of integration.

If the equations (23) are to represent the extremal arcs of an integral I of the form (9), then by Theorem II the functions F and G in (24) must be the solutions for y'' and z'', respectively, of the system of equations

$$H(x, y, z, y', z', y'', z'') = 0, \qquad K(x, y, z, y', z', y'', z'') = 0$$

whose equations of variation are self-adjoint. Hence, there must exist a set of multipliers P, Q_1, Q_2, R such that the equations

$$P(y'' - F) + Q_1(z'' - G) = 0,$$
$$Q_2(y'' - F) + R(z'' - G) = 0$$

can be rendered identical with equations (12) in the previous section. It is evident from the form of (12) that the desired multipliers have $Q_1 = Q_2$ and hence, that the functions H and K may be expressed in the following way:

$$H = Py'' + Qz'' - (PF + QG) = 0,$$
$$K = Qy'' + Rz'' - (QF + RG) = 0. \tag{25}$$

The functions M and N of equations (12) now have the values

$$M = -(PF + QG), \qquad N = -(QF + RG). \tag{26}$$

Necessary and sufficient conditions for the system (25) to be self-adjoint may, therefore, be obtained from Theorem III. Evidently, the first two relations remain as in (16), namely,

$$P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'}. \tag{27}$$

Making the substitution (26) in (16_3), (16_5), (16_4) respectively, one obtains the following system:

$$\begin{aligned} &P_x + P_y y' + P_z z' = -(PF + QG)_{y'}, \\ &Q_x + Q_y y' + Q_z z' = -\tfrac{1}{2}[(PF + QG)_{z'} + (QF + RG)_{y'}], \\ &R_x + R_y y' + R_z z' = -(QF + RG)_{z'}. \end{aligned} \tag{28}$$

With the use of (27) these equations may be written in the form

$$\begin{aligned} &P_x + P_y y' + P_z z' + P_{y'}F + P_{z'}G = -(PF_{y'} + QG_{y'}), \\ &Q_x + Q_y y' + Q_z z' + Q_{y'}F + Q_{z'}G = -\tfrac{1}{2}(PF_{z'} + QG_{z'} + QF_{y'} + RG_{y'}), \\ &R_x + R_y y' + R_z z' + R_{y'}F + R_{z'}G = -(QF_{z'} + RG_{z'}). \end{aligned} \tag{29}$$

If in the last of equations (16) M and N are replaced by their values in (26), we get

$$\begin{aligned} (QF + RG)_y - (PF + QG)_z \equiv \tfrac{1}{2}[&(QF_{y'} + RG_{y'} - PF_{z'} - QG_{z'})_x \\ &+ (QF_{y'} + RG_{y'} - PF_{z'} - QG_{z'})_y y' + (QF_{y'} + RG_{y'} - PF_{z'} - QG_{z'})_z z']. \end{aligned} \tag{30}$$

From the first two of (15), which were shown to be consequences of the first five of relations (16), arise by the substitution (26) the following equations:

$$\begin{aligned} Q_y - P_z &= (Q_x + Q_y y' + Q_z z')_{y'} - (P_x + P_y y' + P_z z')_{z'}, \\ R_y - Q_z &= (R_x + R_y y' + R_z z')_{y'} - (Q_x + Q_y y' + Q_z z')_{z'}, \end{aligned}$$

which by means of (27) and (28) reduce to

$$\begin{aligned} Q_y - P_z = \tfrac{1}{2}[&(P_{y'}F_{z'} + PF_{y'z'} + Q_{y'}G_{z'} + QG_{y'z'}) \\ &- (Q_{y'}F_{y'} + QF_{y'y'} + R_{y'}G_{y'} + RG_{y'y'})], \\ R_y - Q_z = \tfrac{1}{2}[&(P_{z'}F_{z'} + PF_{z'z'} + Q_{z'}G_{z'} + QG_{z'z'}) \\ &- (Q_{z'}F_{y'} + QF_{y'z'} + R_{z'}G_{y'} + RG_{y'z'})]. \end{aligned}$$

The expressions on the left are now replaced by those on the right, respectively, in the first member of (30) after the indicated differentiation is performed. After performing the indicated differentiation in the second member of (30) and employing relations (28), the result readily reduces to

$$\begin{aligned} P[F_z + \tfrac{1}{4}F_{y'}F_{z'} + \tfrac{1}{4}F_{z'}G_{z'} - \tfrac{1}{2}F'_{z'}] + Q[&(G_z - F_y) + \tfrac{1}{4}(G_{z'}^2 - F_{y'}^2) \\ &+ \tfrac{1}{2}(F_{y'} - G_{z'})'] + R[-G_y - \tfrac{1}{4}G_{y'}G_{z'} - \tfrac{1}{4}F_{y'}G_{y'} + \tfrac{1}{2}G'_{y'}] \equiv 0, \end{aligned} \tag{31}$$

where it is understood that

$$F'_{x'} \equiv F_{x'x} + F_{x'y}y' + F_{x'z}z' + F_{x'y'}F + F_{x'z'}G,$$
$$(F_{y'} - G_{z'})' \equiv (F_{y'} - G_{z'})_x + (F_{y'} - G_{z'})_y y' + (F_{y'} - G_{z'})_z z' + (F_{y'} - G_{z'})_{y'}F + (F_{y'} - G_{z'})_{z'}G,$$
$$G'_{y'} \equiv G_{y'x} + G_{y'y}z' + G_{y'z}y' + G_{y'y'}F + G_{y'z'}G.$$

In view of the relations (27), (29), and (31) we have the following theorem:

THEOREM V. *If the solutions $y=y(x)$, $z=z(x)$ of*

$$y'' = F(x,y,z,y',z'), \qquad z'' = G(x,y,z,y',z')$$

are to be the totality of extremal arcs for an integral of the form (9), *then there must exist a set of multipliers P, Q, R which are functions of x, y, z, y', z' such that the equations*

$$H = P(y'' - F) + Q(z'' - G) = 0,$$
$$K = Q(y'' - F) + R(z'' - G) = 0$$

have equations of variation which are self-adjoint along every arc $y=y(x)$, $z=z(x)$. Necessary and sufficient conditions for such multipliers to exist are

$$P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'},$$
$$P_x + P_y y' + P_z z' + P_{y'}F + P_{z'}G = -(PF_{y'} + QG_{y'}),$$
$$Q_x + Q_y y' + Q_z z' + Q_{y'}F + Q_{z'}G = -\tfrac{1}{2}(PF_{z'} + QG_{z'} + QF_{y'} + RG_{y'}),$$
$$R_x + R_y y' + R_z z' + R_{y'}F + R_{z'}G = -(QF_{z'} + RG_{z'}),$$
$$P[F_z + \tfrac{1}{4}F_{y'}F_{z'} + \tfrac{1}{4}F_{z'}G_{z'} - \tfrac{1}{2}F'_{z'}] + Q[(G_z - F_y) + \tfrac{1}{4}(G^2_{z'} - F^2_{y'}) + \tfrac{1}{2}(F_{y'} - G_{z'})'] + R[-G_y - \tfrac{1}{4}G_{y'}G_{z'} - \tfrac{1}{4}F_{y'}G_{y'} + \tfrac{1}{2}G'_{y'}] \equiv 0, \tag{32}$$

which must be identities in x, y, z, y', z'.

As is evident from the preceding discussion no criterion has been found whereby one can determine in advance whether the given functions F and G are solutions for y'' and z'', respectively, of a system of equations

$$H(x,y,z,y',z',y'',z'') = 0, \qquad K(x,y,z,y',z',y'',z'') = 0$$

whose equations of variation are self-adjoint. If, however, the functions F and G are such that conditions (32) can be satisfied then the solutions of (24), $y=y(x)$, $z=z(x)$, represent the extremal arcs for an integral I of the form (9). Some difficulties have been encountered in attempting to satisfy conditions (32) and the writer has not been able to develop completely the theory for the general case. However, the following discussion together with

the examples treated below will illustrate a method of procedure which may be employed with success in special cases.

The theory of partial differential equations assures us of the existence of solutions P, Q, R of the system (29). To find these solutions let us introduce the new variables x, a, b, c, d in place of x, y, z, y', z' by means of the equations

$$\begin{aligned} y &= y(x,a,b,c,d), & y' &= y_x(x,a,b,c,d), \\ z &= z(x,a,b,c,d), & z' &= z_x(x,a,b,c,d), \end{aligned} \tag{33}$$

which have solutions of the form

$$\begin{aligned} a &= A(x,y,z,y',z'), & c &= C(x,y,z,y',z'), \\ b &= B(x,y,z,y',z'), & d &= D(x,y,z,y',z'). \end{aligned} \tag{34}$$

It is readily verified that every set of functions P, Q, R of x, a, b, c, d which satisfy (29), when a, b, c, d are replaced by the expressions (34), must satisfy the equations

$$\begin{aligned} \frac{dP}{dx} &= -(PF_{y'} + QG_{y'}), \\ \frac{dQ}{dx} &= -\tfrac{1}{2}(PF_{z'} + QG_{z'} + QF_{y'} + RG_{y'}), \\ \frac{dR}{dx} &= -(QF_{z'} + RG_{z'}), \end{aligned} \tag{35}$$

and conversely. This follows because each of the functions (34) is a solution of the homogeneous equation of the type

$$A_x + A_y y' + A_z z' + A_{y'} F + A_{z'} G \equiv 0.$$

In the equations (35) the variables x, y, z, y', z' are everywhere to be replaced by x, a, b, c, d by means of the equations (33).

According to the theory of ordinary differential equations, if P_1, Q_1, R_1; P_2, Q_2, R_2; P_3, Q_3, R_3 are any three independent particular solutions of the system (35), then every other solution can be expressed in the form

$$\begin{aligned} C_1P_1 + C_2P_2 + C_3P_3 &= P(x,a,b,c,d), \\ C_1Q_1 + C_2Q_2 + C_3Q_3 &= Q(x,a,b,c,d), \\ C_1R_1 + C_2R_2 + C_3R_3 &= R(x,a,b,c,d), \end{aligned} \tag{36}$$

where C_1, C_2, C_3 are arbitrary functions of a, b, c, d. If C_1, C_2, C_3 are determined in any manner and the functions a, b, c, d are replaced by their

respective values in (34) the resulting values of P, Q, R are solutions of the system (29). Conversely, every solution of the system (29) can be so obtained.

The functions C_1, C_2, C_3 must be determined so that the remaining conditions for self-adjointness of Theorem V, namely (27) and (31), shall be satisfied. The possibility of satisfying these conditions by a suitable choice of the arbitrary functions C_1, C_2, C_3 is not demonstrated in this paper. Such a choice can be made only if the functions F and G satisfy suitable restrictions which do not seem to be easy to characterize explicitly. In the following sections it is shown that for special examples the method can be carried through and the most general integral with a given family of extremals determined.

V. Integrals whose extremals are straight lines

The example of this section is inserted chiefly to show the application of the above theory to the problem which has already been treated by Hamel.*

Consider the equations

$$y'' = 0, \qquad z'' = 0. \tag{37}$$

If these are the solutions for y'' and z'' of two equations

$$H(x,y,z,y',z',y'',z'') = 0, \qquad K(x,y,z,y',z',y'',z'') = 0$$

whose equations of variation are a self-adjoint system, then there exists a set of multipliers P, Q, R such that the functions H and K take the form

$$H = Py'' + Qz'' = 0, \qquad K = Qy'' + Rz'' = 0. \tag{38}$$

The system (32) of Theorem V for this problem becomes the following:

$$\begin{aligned} &P_{z'} = Q_{y'}, \quad Q_{z'} = R_{y'}, \\ &P_z + P_y y' + P_z z' = 0, \\ &Q_z + Q_y y' + Q_z z' = 0, \\ &R_z + R_y y' + R_z z' = 0, \end{aligned} \tag{39}$$

which must be identities in x, y, z, y', z'. The last equation of (32) is identically zero since $F \equiv G \equiv 0$. Differentiate the third of equations (39) with respect to z' and the fourth with respect to y' and subtract. Proceed similarly with the fourth and fifth equations. Then, with the aid of the first two relations of (39), we have

* G. Hamel, *Die Geometrieen im Raume*, Mathematische Annalen, vol. 57, p. 255 ff.

$$P_z = Q_y, \qquad Q_z = R_y. \tag{40}$$

It is evident from the preceding section that, as a preliminary to the solving of the last three equations of the system (39), we must solve the equations

$$y'' = 0, \qquad z'' = 0.$$

By integration we find the equations

$$y' = a_1, \qquad z' = a_2,$$
$$y = a_1 x + a_3, \qquad z = a_2 x + a_4,$$

analogous to the equations (33); where a_1, a_2, a_3, a_4 are constants of integration. The values of a_1, a_2, a_3, a_4 in terms of x, y, z, y', z' found by solving the above equations are

$$a_1 = y', \qquad a_2 = z', \tag{41}$$
$$a_3 = y - y'x, \qquad a_4 = z - z'x,$$

which correspond to the values (34) in the general case.

The equations corresponding to (35) are the following:

$$\frac{dP}{dx} = 0, \quad \frac{dQ}{dx} = 0, \quad \frac{dR}{dx} = 0.$$

Hence, the solutions of the last three equations of (39) are arbitrary functions

$$P(a_1, a_2, a_3, a_4), \qquad Q(a_1, a_2, a_3, a_4), \qquad R(a_1, a_2, a_3, a_4)$$

of the functions a_1, a_2, a_3, a_4.

Thus necessary and sufficient conditions for the system (38) to be self-adjoint are

$$P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'},$$
$$P = P(a_1, a_2, a_3, a_4), \quad Q = Q(a_1, a_2, a_3, a_4), \tag{42}$$
$$R = R(a_1, a_2, a_3, a_4),$$

in which a_1, a_2, a_3, a_4 have the values (41). One can readily show that the functions P, Q, R can be determined so that the first two of relations (42) are satisfied.

To do this let P, Q, R be differentiated with respect to $a_1, \cdots, a_4$, and these in turn differentiated with respect to y' and z'. We have for the first and second of (42) respectively

$$P_2 - P_4 x \equiv Q_1 - Q_3 x,$$
$$Q_2 - Q_4 x \equiv R_1 - R_3 x.$$

Regarding these equations as identities in x we must have

$$P_2 = Q_1, \quad P_4 = Q_3, \quad Q_2 = R_1, \quad Q_4 = R_3. \tag{43}$$

If the first equation is differentiated with respect to a_4 and the second with respect to a_2, we see that the two equations are compatible if

$$Q_{14} = Q_{32}.$$

Similarly, differentiating the last two equations with respect to a_3, a_1, respectively, we arrive at the same condition. Hence, there exist values of P, Q, R which satisfy this system. It follows that the necessary and sufficient conditions (42) for self-adjointness of equations (38) can be satisfied.

If a value of Q is selected satisfying the last equation, then by (43) P is determined to an additive function of a_1, a_3, and R is determined to an additive function of a_2, a_4. In the functions thus determined for P, Q, R let $a_1, \cdots, a_4$ be replaced by their respective values in (41); the resulting functions satisfy the self-adjoint relations (39). Using these values we find a particular solution of the system

$$\begin{aligned} g_{y'y'} &= P(y', z', y - y'x, z - z'x), \\ g_{y'z'} &= Q(y', z', y - y'x, z - z'x), \\ g_{z'z'} &= R(y', z', y - y'x, z - z'x). \end{aligned} \tag{44}$$

Since the conditions for integrability of this system are the first two of relations (39), which also insure that the following integrals are independent of the path of integration, the value of g is given by the double integral

$$g = \int_{y_0'z_0'}^{y'z'} \int_{y_0'z_0'}^{y'z'} Pdy'dy' + 2Qdy'dz' + Rdz'dz'.$$

Corresponding to the line integrals M and L of §III, we have

$$g_{y'} = \int_{y_0'z_0'}^{y'z'} Pdy' + Qdz', \qquad g_{z'} = \int_{y_0'z_0'}^{y'z'} Qdy' + Rdz'.$$

Applying relations (40), which were shown to be consequences of (39), to the last two integrals we get

$$g_{y'z} = g_{z'y}. \tag{45}$$

By means of this relation and the fact that the functions M and N of (26) are each identically zero, equations (21) of §III become

$$\begin{aligned} c_y - b_z &= 0, \\ a_z - c_z &= -g_z + g_{z'x} + g_{y'z}y' + g_{z'z}z', \\ b_x - a_y &= g_y - g_{y'x} - g_{y'y}y' - g_{z'y}z'. \end{aligned} \tag{46}$$

It is readily verified that this system satisfies the required conditions (22) which insure the existence of solutions for a, b, c.

In this particular case it is easy to show that, on account of relations (40), the second members of the last two equations vanish identically. By means of the integrals given above for g and $g_{y'}$ we have

$$g_y - g_{y'x} - g_{y'y}y' - g_{y'z}z' = \int_{y_0'z_0'}^{y'z'} \int_{y_0'z_0'}^{y'z'} P_y dy'dy' + 2Q_y dy'dz' + R_y dz'dz'$$

$$- \int_{y_0'z_0'}^{y'z'} P_x dy' + Q_x dz' - y' \int_{y_0'z_0'}^{y'z'} P_y dy' + Q_y dz' - z' \int_{y_0'z_0'}^{y'z'} P_x dy' + Q_x dz'.$$

If in the second integral the values of P_x and Q_x obtained from (39) are inserted, we have

$$- \int_{y_0'z_0'}^{y'z'} P_x dy' + Q_x dz' = \int_{y_0'z_0'}^{y'z'} y'(P_y dy' + Q_y dz') + z'(P_x dy' + Q_x dz')$$

$$= y' \int_{y_0'z_0'}^{y'z'} P_y dy' + Q_y dz' - \int_{y_0'z_0'}^{y'z'} \int_{y_0'z_0'}^{y'z'} P_y dy'dy' + Q_y dy'dz'$$

$$- z' \int_{y_0'z_0'}^{y'z'} P_x dy' + Q_x dz' - \int_{y_0'z_0'}^{y'z'} \int_{y_0'z_0'}^{y'z'} P_x dy'dz' + Q_x dz'dz'.$$

When this value of the integral on the left is substituted in the previous equation the result is identically zero in view of the relations (40). The second member of (46_2) may be treated similarly. Hence, we have the conditions

$$c_y - b_z = 0, \qquad a_z - c_x = 0, \qquad b_x - a_y = 0.$$

Therefore, the integrand f of the integral whose extremals are straight lines is of the form

$$f = g(x, y, z, y', z') + \frac{d}{dx} t(x, y, z),$$

where g is a particular solution of the system (44) and t is an arbitrary function of x, y, z.

VI. Integrals whose extremals are semi-circles orthogonal to the xy-plane

Consider the four-parameter family of circles

$$\begin{aligned} x &= a + b \cos d \cos u, \\ y &= c - b \sin d \cos u, \\ z &= b \sin u, \end{aligned} \tag{47}$$

where u is the angle which the radius vector makes with the line of intersection of the plane of the circle and the xy-plane, and d is the angle which this line of intersection makes with the positive x-axis.

The differential equations of these curves found by eliminating the parameter u and the four constants a, b, c, d are

$$y'' = 0, \quad z'' = -\frac{1 + y'^2 + z'^2}{z}. \tag{48}$$

When the second members of these equations are substituted for F and G in the system (32) and the independent variables x, y, z, y', z' replaced by x, a, b, c, d this system takes the form

$$\begin{aligned} &P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'}, \\ &\frac{dP}{dx} = \frac{2Qy'}{z}, \quad \frac{dQ}{dx} = \frac{Qz'}{z} + \frac{Ry'}{z}, \quad \frac{dR}{dx} = \frac{2Rz'}{z}. \end{aligned} \tag{49}$$

The coefficients in the last equation of (32) for this problem all vanish identically on account of the form of F and G.

The integrals of equations (48) are the curves (47). When the variable u is eliminated from the equations (47) and the derivatives y', z' with respect to x are calculated we find

$$\begin{aligned} &y = c + (a - x)\tan d, \quad z = \sec d(b^2\cos^2 d - (x - a)^2)^{1/2}, \\ &y' = -\tan d, \qquad\qquad z' = \frac{\sec d(a - x)}{(b^2\cos^2 d - (x - a)^2)^{1/2}}. \end{aligned}$$

When these are solved for a, b, c, d the resulting functions

$$\begin{aligned} &a = x + \frac{zz'}{1 + y'^2}, \qquad b^2 = z^2 + \frac{z^2z'^2}{1 + y'^2}, \\ &c = y + \frac{zy'z'}{1 + y'^2}, \qquad \tan d = -y' \end{aligned} \tag{50}$$

are four independent solutions of the homogeneous equation

$$A_x + A_y y' + A_z z' + A_{y'}0 - A_{z'}\frac{1 + y'^2 + z'^2}{z} \equiv 0$$

which characterizes the first integral of equations (48). But if (50) are solutions of this equation then also is any function of them a solution and in particular the following:

$$\begin{aligned} &a_1 = c + a\tan d, \qquad a_2 = a - c\tan d, \\ &a_3 = b^2\sec^2 d, \qquad\quad a_4 = -\tan d \end{aligned}$$

which are selected because of their simplicity in terms of the variables x, y, z, y', z'. In terms of these variables they have the values

$$(51)\qquad \begin{aligned} a_1 &= y - y'x, & a_2 &= x + yy' + zz', \\ a_3 &= z(1 + y'^2 + z'^2), & a_4 &= y'. \end{aligned}$$

It is evident from the form of equations (49) that the last one can be integrated directly and, hence, also the second and the first. The general solution of the system is

$$(52)\qquad \begin{aligned} P &= \alpha + 2\beta y + \gamma y^2, \\ Q &= \beta z + \gamma yz, \\ R &= \gamma z^2, \end{aligned}$$

where α, β, γ are constants of integration and arbitrary functions of a_1, a_2, a_3, a_4.

With the values of F and G given in (48) it is readily verified, as has been mentioned above, that the coefficients of P, Q, R in the last equation of (32) are each identically zero. This is not a characteristic property of the functions F and G for every such problem of the calculus of variations, though it holds for this particular case.

It remains now to show that there exist values of α, β, γ such that the remaining conditions for self-adjointness, namely

$$P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'},$$

are satisfied. Imposing these conditions on the second members in (52), we have

$$\begin{aligned} y^2\gamma_{z'} + 2y\beta_{z'} + \alpha_{z'} &\equiv yz\gamma_{y'} + z\beta_{y'}, \\ y\,\gamma_{z'} + \beta_{z'} &\equiv z\gamma_{y'}, \end{aligned}$$

which after a simple recombination are equivalent to the following:

$$(53)\qquad \begin{aligned} z\gamma_{y'} - y\gamma_{z'} &= \beta_{z'}, \\ z\beta_{y'} - y\beta_{z'} &= \alpha_{z'}. \end{aligned}$$

After differentiating α, β, γ with respect to a_1, a_2, a_3, a_4 and these in turn, as given in (51), with respect to y' and z', equations (53) become

$$(54)\qquad \begin{aligned} -x\gamma_1 + 2z(y'z - yz')\gamma_3 + \gamma_4 - \beta_2 - 2zz'\beta_3 &= 0, \\ -x\beta_1 + 2z(y'z - yz')\beta_3 + \beta_4 - \alpha_2 - 2zz'\alpha_3 &= 0, \end{aligned}$$

in which the subscripts 1, 2, 3, 4 denote differentiation with respect to a_1, a_2, a_3, a_4 respectively. If equations (51) are solved for y, z, y', z' and their values, namely

$$y = a_1 + a_4 x, \quad z = \left(\frac{a_3 - [(1 + a_4^2)x + (a_1a_4 - a_2)]^2}{1 + a_4^2}\right)^{1/2},$$

$$y' = a_4, \qquad z' = \left(\frac{(1 + a_4^2)[a_2 - a_1a_4 - (1 + a_4^2)x]^2}{a_3 - [(1 + a_4^2) + (a_1a_4 - a_2)]^2}\right)^{1/2},$$

are substituted in equations (54), the resulting equations must be identities in $x, a_1, \cdots, a_4$. This substitution is performed in the first equation of (54) and the coefficients collected according to the powers of x. The coefficient of x^2 vanishes; that of x and the constant term give the first two of the following conditions. Similar results from the second equation of (54) follow by symmetry and give the last two conditions:

$$\begin{aligned}
&2(a_1 + a_2a_4)\gamma_3 + 2(1 + a_4^2)\beta_3 - \gamma_1 = 0,\\
&2[a_4a_3 - (a_1 + a_2a_4)(a_2 - a_1a_4)]\gamma_3 - 2(a_2 - a_1a_4)(1 + a_4^2)\beta_3\\
&\qquad + (1 + a_4^2)\gamma_4 - (1 + a_4^2)\beta_2 = 0,\\
&2(a_1 + a_2a_4)\beta_3 + 2(1 + a_4^2)\alpha_3 - \beta_1 = 0,\\
&2[a_4a_3 - (a_1 + a_2a_4)(a_2 - a_1a_4)]\beta_3 - 2(a_2 - a_1a_4)(1 + a_4^2)\alpha_3\\
&\qquad + (1 + a_4^2)\beta_4 - (1 + a_4^2)\alpha_2 = 0.
\end{aligned} \tag{55}$$

The first two of equations (55) can be solved for β_2 and β_3. One obtains

$$\begin{aligned}
(1 + a_4^2)\beta_2 &= -(a_2 - a_1a_4)\gamma_1 + 2a_3a_4\gamma_3 + (1 + a_4^2)\gamma_4,\\
(1 + a_4^2)\beta_3 &= \tfrac{1}{2}\gamma_1 - (a_1 + a_2a_4)\gamma_3.
\end{aligned} \tag{56}$$

If these equations are differentiated with respect to a_3 and a_4 respectively, then

$$\begin{aligned}
(1 + a_4^2)\beta_{23} &= -(a_2 - a_1a_4)\gamma_{13} + 2a_3a_4\gamma_{33} + 2a_4\gamma_3 + (1 + a_4^2)\gamma_{43},\\
(1 + a_4^2)\beta_{32} &= \tfrac{1}{2}\gamma_{12} - (a_1 + a_2a_4)\gamma_{32} - a_4\gamma_3.
\end{aligned} \tag{57}$$

If the equations (56) are consistent the second members of (57) must be identical. Hence, we have the condition

$$\begin{aligned}
3a_4\gamma_3 = (a_2 - a_1a_4)\gamma_{13} - (a_1 + a_2a_4)\gamma_{32} - 2a_3a_4\gamma_{33}\\
- (1 + a_4^2)\gamma_{43} + \tfrac{1}{2}\gamma_{12}
\end{aligned} \tag{58}$$

that must be satisfied by the function γ. A pair of equations similar to (56) for α and β is obtained from the last two equations of (55). They are, in fact, the system (56) with β, γ replaced by α, β respectively. These new equations are compatible only if β satisfies the equation obtained from (58) by replacing γ by β. If in this equation for β its derivatives with respect

to a_1, a_2, a_3, a_4 are replaced by their respective values obtained from (56) and (57) the result again reduces to the condition (58).

Therefore, equations (53) are consistent and there exist values of the functions α, β, γ such that the conditions

$$P_{z'} = Q_{y'}, \quad Q_{z'} = R_{y'}$$

are satisfied provided that γ satisfies equation (58).

THEOREM VI. *The most general integral*

$$I = \int_{x_1}^{x_2} f(x, y, z, y', z')dx$$

whose extremals are the family of circles (47) *satisfying the equation* (48) *is determined as follows:*

For an arbitrary solution $\gamma(a_1, a_2, a_3, a_4)$ *of the equation* (58) *the equations* (56) *are compatible and determine* $\beta(a_1, a_2, a_3, a_4)$ *to an arbitrary additive function of* a_1, a_4. *Equations* (56) *with* γ *replaced by this function* β *and* β *replaced by* α, *are again compatible and define* $\alpha(a_1, a_2, a_3, a_4)$ *to an arbitrary additive function of* a_1, a_4. *The functions*

$$\begin{aligned} P(x, y, z, y', z') &= \alpha + 2\beta y + \gamma y^2, \\ Q(x, y, z, y', z') &= \beta z + \gamma y^2, \\ R(x, y, z, y', z') &= \gamma z^2 \end{aligned}$$

so determined satisfy the conditions (49) *which are necessary and sufficient for the equations*

$$H = Py'' + Q\left(z'' + \frac{1 + y'^2 + z'^2}{z}\right) = 0,$$

$$K = Qy'' + R\left(z'' + \frac{1 + y'^2 + z'^2}{z}\right) = 0$$

to have equations of variation which are self-adjoint. By the process of §III *the integrand function* $f(x, y, z, y', z')$ *is then determined to an additive derivative of an arbitrary function of* x, y, z. *The arbitrary elements are, therefore, the choice of an integral of the partial differential equation* (58) *of the second order, the choice of two arbitrary additive functions of* a_1 *and* a_4, *and the choice of the derivative of an arbitrary additive function of* (x, y, z).

VII. Integrals whose extremals are catenaries orthogonal to the xy-plane

Consider the four-parameter family of catenaries

$$
\begin{aligned}
x &= a + c\cos\theta\cosh^{-1}\left(\frac{s^2+c^2}{c}\right)^{1/2},\\
y &= b - c\sin\theta\cosh^{-1}\left(\frac{s^2+c^2}{c}\right)^{1/2},\\
z &= (s^2+c^2)^{1/2},
\end{aligned}
\tag{59}
$$

where s is the arc length of the catenary measured from the lowest point and θ is the angle which the line of intersection of the plane of the catenary and of the xy-plane makes with the positive x-axis.

The differential equations of the curves (1) found by eliminating the parameter s and the four constants a, b, c, θ are

$$
y'' = 0, \quad z'' = \frac{1+y'^2+z'^2}{z}. \tag{60}
$$

If these values of y'' and z'' are substituted for F and G respectively in the system (32), §IV, the latter take the following form:

$$
\begin{aligned}
&P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'},\\
&\frac{dP}{dx} = -Q\frac{2y'}{z}, \quad \frac{dQ}{dx} = -\left(Q\frac{z'}{z} + R\frac{y'}{z}\right), \quad \frac{dR}{dx} = -R\frac{2z'}{z},\\
&Q(1+y'^2) + Ry'z' = 0.
\end{aligned}
\tag{61}
$$

The integrals of (60) are the curves (59). When the parameter s is eliminated from the equations (59) and the derivatives y' and z' are found, one obtains

$$
\begin{aligned}
&y = b - (x-a)\tan\theta, \quad z = c\cosh\left(\frac{x-a}{c\cos\theta}\right),\\
&y' = -\tan\theta, \qquad z' = \sec\theta\sinh\left(\frac{x-a}{c\cos\theta}\right).
\end{aligned}
\tag{62}
$$

If the substitutions

$$
a_1 = -\tan\theta, \qquad a_2 = b, \qquad a_3 = c, \qquad a_4 = a
$$

are made equations (62) become

$$
\begin{aligned}
&y = a_2 - a_1a_4 + a_1x, \quad z = a_3\cosh\left(\frac{(x-a_4)(1+a_1^2)^{1/2}}{a_3}\right),\\
&y' = a_1, \qquad\qquad z' = (1+a_1^2)^{1/2}\sinh\frac{(x-a_4)(1+a_1^2)^{1/2}}{a_3}.
\end{aligned}
\tag{63}
$$

When these equations are solved for a_1, a_2, a_3, a_4 the resulting functions

$$
\begin{aligned}
&a_1 = y', \quad a_2 = y - \frac{y'_z}{(1+y'^2+z'^2)^{1/2}}\cosh^{-1}\left(\frac{1+y'^2+z'^2}{1+y'^2}\right)^{1/2},\\
&a_3 = \frac{z(1+y'^2)^{1/2}}{(1+y'^2+z'^2)^{1/2}},\\
&a_4 = x - \frac{z}{(1+y'^2+z'^2)^{1/2}}\cosh^{-1}\left(\frac{1+y'^2+z'^2}{1+y'^2}\right)^{1/2}
\end{aligned}
\tag{64}
$$

are four independent solutions of the homogeneous equation

$$
A_x + A_y y' + A_z z' + A_{y'}0 + A_{z'}\frac{1+y'^2+z'^2}{z} = 0,
$$

which characterizes the first integral of equations (60).

The fifth equation of the system (61) can be integrated directly and then also the fourth and the third. They yield the following system:

$$
\begin{aligned}
&P = \alpha - \frac{2\beta y'z'}{z(1+y'^2)} - \frac{\gamma y'^2}{z^2(1+y'^2)},\\
&Q = \frac{\beta}{z} - \frac{\gamma y'z'}{z^2(1+y'^2)},\\
&R = \frac{\gamma}{z^2};
\end{aligned}
\tag{65}
$$

where α, β, γ are arbitrary functions of $a_1, \cdots, a_4$. The above values of P, Q, R must satisfy the last of equations (61). We thus find that β must be identically zero. The system (65) now becomes

$$
\begin{aligned}
&P = \alpha - \frac{\gamma y'^2}{z^2(1+y'^2)},\\
&Q = -\frac{\gamma y'z'}{z^2(1+y'^2)},\\
&R = \frac{\gamma}{z^2}.
\end{aligned}
\tag{66}
$$

If these functions are to satisfy the first two relations of (61) we must have the following conditions satisfied by α and γ:

$$\frac{z'(y'^2-1)\gamma}{1+y'^2} - y'z'\gamma_{y'} + y'^2\gamma_{z'} = z^2(1+y'^2)\alpha_{z'},$$

$$y'\gamma + (1+y'^2)\gamma_{y'} + y'z'\gamma_{z'} = 0,$$

which shall be identities in $x, a_1, \cdots, a_4$. These equations are equivalent to the system

$$\begin{aligned} z'(2y'^2-1)\gamma + y'^2(1+y'^2+z'^2)\gamma_{z'} &= z^2(1+y'^2)^2\alpha_{z'}, \\ y'\gamma + (1+y'^2)\gamma_{z'} + y'z'\gamma_{z'} &= 0. \end{aligned} \tag{67}$$

Since α, γ are functions of $a_1, \cdots, a_4$ they may be differentiated with respect to $a_1, \cdots, a_4$ and these functions in turn differentiated with respect to y' and z'. Thus equations (67) give the following relations:

$$\begin{aligned} &\gamma y' + \gamma_1(1+y'^2) - \gamma_2\frac{z\cosh^{-1}u}{(1+y'^2+z'^2)^{1/2}} + \gamma_3\cdot 0 + \gamma_4\frac{zy'\cosh^{-1}u}{(1+y'^2+z'^2)^{1/2}} = 0, \\ &\gamma z'(2y'^2-1) + \gamma_2\left(\frac{zy'^3z'\cosh^{-1}u}{(1+y'^2+z'^2)^{1/2}} - zy'^3\right) - \gamma_3\frac{zy'^2z'(1+y'^2)^{1/2}}{(1+y'^2+z'^2)^{1/2}} \\ &\qquad + \gamma_4\frac{zy'^2z'\cosh^{-1}u - zy'^2}{(1+y'^2+z'^2)^{1/2}} \\ &\qquad = \alpha_2\left[\frac{z^3y'z'(1+y'^2)^2\cosh^{-1}u}{(1+y'^2+z'^2)^{3/2}} - \frac{y'z^3(1+y'^2)^2}{(1+y'^2+z'^2)}\right] \\ &\quad - \alpha_3\frac{z^3z'(1+y'^2)^{5/2}}{(1+y'^2+z'^2)^{3/2}} + \alpha_4\left[\frac{z^3z'(1+y'^2)^2\cosh^{-1}u}{(1+y'^2+z'^2)^{3/2}} - \frac{z^3(1+y'^2)^2}{(1+y'^2+z'^2)}\right]; \end{aligned} \tag{68}$$

where the subscripts 1, 2, 3, 4 indicate differentiation with respect to a_1, a_2, a_3, a_4 respectively and

$$u = \left(\frac{1+y'^2+z'^2}{1+y'^2}\right)^{1/2}.$$

If the values given in (63) for y, z, y', z' respectively are substituted in (68), one obtains

$$\begin{aligned} &a_1\gamma + (1+a_1^2)\gamma_1 - (x-a_4)\gamma_2 + a_1(x-a_4)\gamma_4 = 0, \\ &\gamma(1+a_1^2)^{1/2}(2a_1^2-1)\sinh\omega + \gamma_2a_1^3[(x-a_4)(1+a_1^2)^{1/2}\sinh\omega - a_3\cosh\omega] \\ &\quad - \gamma_3a_1^2a_3(1+a_1^2)^{1/2}\sinh\omega + \gamma_4a_1^3[(x-a_4)(1+a_1^2)^{1/2}\sinh\omega - a_3\cosh\omega] \\ &= \alpha_2a_1a_3^2(1+a_1^2)[(x-a_4)(1+a_1^2)^{1/2}\sinh\omega - a_3\cosh\omega] \\ &\quad - \alpha_3a_3^3(1+a_1^2)^{3/2}\sinh\omega + \alpha_4a_3^2(1+a_1^2)[(x-a_4)(1+a_1^2)^{1/2}\sinh\omega \\ &\qquad - a_3\cosh\omega] \end{aligned} \tag{69}$$

in which $$\omega = \frac{(x - a_4)(1 + a_1^2)^{1/2}}{a_3}.$$

The last equation may be written in the form

$$\begin{aligned}(69') \quad & [(1 + a_1^2)^{1/2}(2a_1^2 - 1)\gamma - a_1^3 a_4(1 + a_1^2)^{1/2}\gamma_2 - a_1^2 a_3(1 + a_1^2)^{1/2}\gamma_3 \\ & - a_1^2 a_4(1 + a_1^2)^{1/2}\gamma_4 + a_1 a_3^2 a_4(1 + a_1^2)^{3/2}\alpha_2 + a_3^3(1 + a_1^2)^{3/2}\alpha_3 \\ & + a_3^2 a_4(1 + a_1^2)^{3/2}\alpha_4] + [a_1^3(1 + a_1^2)^{1/2}\gamma_2 + a_1^2(1 + a_1^2)^{1/2}\gamma_4 \\ & - a_1 a_3^2(1 + a_1^2)^{3/2}\alpha_2 - a_3^2(1 + a_1^2)^{3/2}\alpha_4]x - [a_1^3 a_3\gamma_2 + a_1^2 a_3\gamma_4 \\ & - a_1 a_3^3(1 + a_1^2)\alpha_2 - a_3^3(1 + a_1^2)\alpha_4] \coth \omega = 0.\end{aligned}$$

These equations must be identities in x. Hence, from the first of (69), we have

$$a_1\gamma_4 - \gamma_2 = 0, \quad a_1\gamma + (1 + a_1^2)\gamma_1 + a_4\gamma_2 - a_1 a_4\gamma_4 = 0,$$

which are equivalent to the system

$$(70) \qquad a_1\gamma_4 - \gamma_2 = 0, \qquad a_1\gamma + (1 + a_1^2)\gamma_1 = 0.$$

Equation (69′) upon equating the respective coefficients to zero gives

$$\begin{aligned} & a_1^2(a_1\gamma_2 + \gamma_4) = a_3^2(1 + a_1^2)(a_1\alpha_2 + \alpha_4), \\ & (2a_1^2 - 1)\gamma - a_1^3 a_4\gamma_2 - a_1^2 a_3\gamma_3 - a_1^2 a_4\gamma_4 \\ & \qquad = - a_3^2(1 + a_1^2)(a_1 a_4\alpha_2 + a_3\alpha_3 + a_4\alpha_4). \end{aligned}$$

The latter are equivalent to the following relations:

$$(71) \qquad \begin{aligned} & a_1^2(a_1\gamma_2 + \gamma_4) = a_3^2(1 + a_1^2)(a_1\alpha_2 + \alpha_4), \\ & (2a_1^2 - 1)\gamma - a_1^2 a_3\gamma_3 = - a_3^3(1 + a_1^2)\alpha_3. \end{aligned}$$

When equations (71) are solved for α_3 and α_4 with the aid of the first of relations (70) one obtains

$$\alpha_4 = \frac{a_1^2}{a_3^2}\gamma_4 - a_1\alpha_2,$$

$$\alpha_3 = \frac{-1}{a_3^3(1 + a_1^2)}[(2a_1^2 - 1)\gamma - a_1^2 a_3\gamma_3].$$

If these equations are compatible the derivatives of the expressions on the right with respect to a_3 and a_4 respectively, after α_{23} in the first expression is replaced by its value obtained from the second, must be identically equal. With the aid of (70_1) the result reduces to the condition

$$(72) \qquad (1 + a_1^2)\gamma_4 = 0.$$

Hence, by the first equation of (70) γ_2 must also be identically zero. Thus γ

must be a function of a_1 and a_3 alone such that it satisfies the second condition of (70). This will determine γ to an additive function of a_3. From the above restriction on γ and the first relation of (71) we have

$$a_1\alpha_2 + \alpha_4 = 0$$

which together with the second equation of (71) will determine α to an additive function of a_1.

It is now evident that equations (67) are compatible and that there exist values of the functions γ and α such that the relations

$$P_{z'} = Q_{y'}, \qquad Q_{z'} = R_{y'}$$

are satisfied provided that γ is a function of a_1, a_3 alone and satisfies the second equation of (70).

THEOREM VII. *The most general integral*

$$I = \int_{x_1}^{x_2} f(x, y, z, y', z')dx$$

whose extremals are the family of catenaries (59) *satisfying equations* (60) *is determined as follows: For an arbitrary solution* $\gamma(a_1, a_3)$ *of the equation* (70_1) *the equations* (71) *are compatible and determine* α *to an arbitrary additive function of* a_1. *The functions*

$$P = \alpha - \frac{\gamma y'^2}{z^2(1 + y'^2)},$$

$$Q = -\frac{\gamma y'z'}{z^2(1 + y'^2)},$$

$$R = \frac{\gamma}{z^2}$$

thus determined satisfy the conditions (61) *which are necessary and sufficient for the equations*

$$H = Py'' + Q\left(z'' - \frac{1 + y'^2 + z'^2}{z}\right) = 0,$$

$$K = Qy'' + R\left(z'' - \frac{1 + y'^2 + z'^2}{z}\right) = 0$$

to have equations of variation which are self-adjoint. By the method of §III *the integrand function* $f(x, y, z, y', z')$ *is then determined to the derivative of an arbitrary function of* x, y, z.

UNIVERSITY OF OREGON,
EUGENE, ORE.

TYPES OF MOTION OF THE GYROSCOPE*

BY

A. H. COPELAND

1. **Introduction.** We can display, by a simple graphical representation, the complete history of all the motions of all gyroscopes which are subject to the following restrictions:† each gyroscope is acted upon by gravity, by a constraint which keeps one point on its axis of symmetry fixed in space, and by no other forces. Furthermore, with the aid of Osgood's intrinsic equations,‡ we shall be able to include in this history a discussion of some new intrinsic properties of the cone traced in space by the axis of symmetry of such a gyroscope. This cone intersects the unit sphere whose center is at the fixed point, in a curve which we shall call Γ. We shall describe the properties of this cone in terms of the geodesic curvature, k, of Γ. The advantage in Osgood's equations lies in the fact that they yield a simple expression for k. We shall also be able to obtain additional properties of the motion from an analysis of the curves Γ passing near the poles (the north pole being at the top of the unit sphere). By means of these methods, we shall list the properties of the motion more fully than has hitherto been done, exhibit their dependence upon the initial conditions, and show how one type of motion changes into another as the initial conditions are varied continuously.

2. **Equations of the motion.** In the case of the gyroscope which we are considering, Osgood's equations take the following form:§

$$(1)\qquad \begin{aligned} &(a) && Avv' = Mgh(\sin\theta)\theta', \\ &(b) && Akv^2 - Crv = -Mgh(\sin^2\theta)\psi', \\ &(c) && C\dot{r} = 0, \end{aligned}$$

---

* Presented to the Society, October 29, 1927; received by the editors December 30, 1927. This paper was written in connection with the author's doctorate.

† For discussions of the motion of such a gyroscope, see, for example Klein und Sommerfeld, *Theorie des Kreisels*, Heft II; Appell, *Traité de la Mécanique Rationnelle*, Tome II, pp. 193–209; Encyklopädie der Mathematischen Wissenschaften, Band IV_1, pp. 619–639; or Lamb, *Higher Mechanics*, Chap. VIII.

‡ Cf. W. F. Osgood, *On the gyroscope*, these Transactions, vol. 23 (1922), pp. 240–264.

§ Cf. Osgood, loc. cit., p. 247, formula (7) and p. 256, formula (1). The term Crv of equation (1b) of this paper is preceded by a minus sign instead of the plus sign in Osgood's article. This change of sign is due to the fact that we have chosen a right-handed instead of a left-handed system of coördinates.

where C is the moment of inertia of the gyroscope about its axis of symmetry; A, the moment of inertia about a perpendicular axis through the fixed point; M, the mass of the gyroscope; h, the distance of the center of mass from the fixed point; g, the acceleration of gravity; and r, the spin of the gyroscope about its axis. Likewise, θ and ψ denote, respectively, the colatitude and longitude of a point, ξ, which is a trace of the axis of the gyroscope on the surface of the unit sphere;* k denotes the geodesic curvature of Γ at the point ξ; and v, the velocity of this point. The primes denote differentiation with respect to the arc length, s, of Γ, and the dot denotes differention with respect to the time, t. Equations (1) are equivalent to the vector equation $\dot{\sigma}=\mu$, where σ is the moment of momentum vector taken with respect to the fixed point and μ is the resultant moment of the external forces with respect to the fixed point. Equation (1c) implies that r is a constant. This constant is positive. Equation (1a) has the following integral:

$$v^2 = v_0^2 + a(u_0 - u), \tag{2}$$

where $a=2Mgh/A$ and $u=\cos\theta$, and where v_0 and u_0 are respectively the values of v and u at some time, t_0.

In addition to equations (1) and (2), we have the following equations:†

$$v^2 = \dot{\theta}^2 + (\sin^2\theta)\dot{\psi}^2, \tag{3}$$

$$(1 - u_0^2)\psi_0^2 + \gamma(u_0 - u) = (1 - u^2)\psi \tag{4}$$

where $\gamma=Cr/A$, and

$$\dot{u}^2 = (1 - u^2)[v_0^2 + a(u_0 - u)] - [(1 - u_0^2)\psi_0 + \gamma(u_0 - u)]^2 = f(u). \tag{5}$$

Equations (4) and (5) define the motion of the gyroscope. If u_0 is a root of $f(u)$, then a solution of these equations is $u\equiv u_0$, $\psi\equiv\psi_0$. It is a familiar fact that if u_0 is a double root of $f(u)$, then this solution satisfies equations $\dot{\sigma}=\mu$, but that otherwise the solution is extraneous.‡ When $u\equiv u_0$ and $\psi\equiv\psi_0$,

* The point ξ is that one of the two traces of the axis on the sphere such that an observer situated at the fixed point and looking at ξ sees the gyroscope rotate in a clockwise sense.

† Equation (3) is a consequence of the fact that the point ξ always lies in the surface of the unit sphere. (Cf. Osgood, loc. cit., p. 247.) Equation (4) is a consequence of the fact that the vector μ is always horizontal and hence the vertical component of the vector σ is a constant. Equation (5) is obtained by combining equations (2), (3), and (4). (For equations (4) and (5), cf. Osgood, loc. cit., p. 257, equation (5); also Appell, loc. cit., Tome II, p. 196, equations 50, 51.)

‡ The intrinsic equations afford the following simple proof of this fact. If $u\equiv u_0$, then by equation (2), v is constant, and hence it follows from equation (1b) that k must be constant. Moreover the value k_0 of k obtained by solving this equation, must be equal to the geodesic curvature k_c of the parallel of latitude, u_0. O. D. Kellogg (*Curvature and the top*, these Transactions, vol. 25 (1923), p. 518, formula 29) has obtained the equation

the gyroscope is said to execute steady precession.* Without loss of generality, we can assume that u_0 is a root of $f(u)$. To see this let us assume that u_0 is not a root of $f(u)$. Then, since $f(u_0) \geqq 0$ and $-1 \leqq u_0 \leqq +1$ for real motion, it follows that $f(u_0) > 0$. Furthermore, since $f(\pm 1) \leqq 0$, then $-1 < u_0 < +1$ and there exist at least two distinct roots of $f(u)$ in the interval $-1 \leqq u \leqq +1$. Let u_1 be one of these roots. Since $f(u)$ is a cubic, u_1 may be so chosen that it is a single root and that no other root of $f(u)$ lies between u_0 and u_1.† Then we may combine the equations $v^2 = v_0^2 + a(u_0 - u)$ and $v_1^2 = v_0^2 + a(u_0 - u_1)$ and obtain the equation

$$v^2 = v_1^2 + a(u_1 - u).$$

Likewise we may combine the equations $(1-u^2)\psi = (1-u_0^2)\psi_0^2 + \gamma(u_0 - u)$ and $(1-u_1^2)\psi_1 = (1-u_0^2)\psi_0 + \gamma(u_0 - u_1)$ and obtain the equation

$$(1 - u^2)\psi = (1 - u_1^2)\psi_1 + \gamma(u_1 - u).$$

Hence we may write

$$f(u) = (1 - u^2)[v_1^2 + a(u_1 - u)] - [(1 - u_1^2)\psi_1 + \gamma(u_1 - u)]^2$$

where $-1 \leqq u_1 \leqq +1$. Thus we may take the cosine of the colatitude of the initial point to be u_1 instead of u_0. We can now let u_0 play the role of u_1, that is, we can assume that u_0 is a root of $f(u)$.

From this point on we shall assume that $f(u_0) = 0$. There is a second root, u_1, of $f(u)$ in the interval $-1 \leqq u_1 \leqq +1$, and a third root $u_2 \geqq +1$. It is a familiar fact that Γ lies completely in the zone bounded by the parallels of latitude $u = u_0$ and $u = u_1$ and that, in general, it oscillates between points of tangency to u_0 and to u_1.

3. **Graphical representation of gyroscopic motion.** Equations (4) and (5) show that the type of motion is completely determined by the values

---

$$k_0 - k_c = -\frac{1 - u_0}{v_0^2}\dot{\psi}_0\left(u_0\dot{\psi}_0^2 - \gamma\dot{\psi}_0 + \frac{1}{2}a\right).$$

But we have also the equation

$$g(u_0) = 2(1 - u_0^2)[u_0\dot{\psi}_0^2 - \gamma\dot{\psi}_0 + \tfrac{1}{2}a],$$

and hence we may write

$$k_0 - k_c = \frac{\dot{\psi}_0}{2v_0^2}\, g(u_0).$$

It follows that $k_0 = k_c$ only when $g(u_0)$ or $\dot{\psi}_0$ vanishes. But $t = u_1$ is the only solution of $g(u) = 0$ for which $-1 < u < +1$, and hence $g(u_0) = 0$ only when $u_0 = u_1$. If the factor $\dot{\psi}_0$ vanishes then, since $\dot{u} \equiv 0$, $v_0 = 0$ and hence $v \equiv 0$. It follows that the moment of momentum vector, σ, is constant and therefore the moment μ is zero. But this is only possible if $u_0 = \pm 1$ and if u_0 is a double root of $f(u)$. (Cf. p. 745.)

* Cf. Lamb, loc. cit., p. 130.

† If u_1 is a double root and initially $u = u_1$, then we have seen that the solution is $u \equiv u_1$. This solution cannot be equivalent to one for which $u = u_0$ initially.

of the parameters u_0, ψ_0, a and γ.* We shall see that a given situation which characterizes a type of motion—for example, a change in the sign in the longitudinal velocity—occurs every time Γ crosses a certain parallel of latitude, u. If the parameters u_0, a, and γ are held constant and u and ψ_0 are allowed to vary, then the locus of points for which a given situation occurs, is a curve in the u, ψ_0-plane. These curves are shown in Figs. I to XI. The loci corresponding to points of tangency of Γ to the parallels of latitude, u_0 and u_1, are respectively the straight line $u=u_0$ and the curve $u=u_1$. Since Γ lies entirely within the zone bounded by the parallels of latitude u_0 and u_1, then in a u, ψ_0-diagram, the points which correspond to real motoin must lie in the region between the line $u=u_0$ and the curve $u=u_1$. In Figs. I to XI, this region is indicated by the shading.

The straight line $u=u_L$ and the parabolas $u=u_i$ and $u=u_k$ are loci which correspond respectively to points at which the longitudinal velocity, ψ, the geodesic curvature, k, and the derivative, k', of k with respect to the arc length, s, of Γ, vanish. By studying these loci, we can determine whether Γ has waves, loops, or cusps, whether or not it has points of inflection, and whether or not its curvature is monotone between points of tangency to the parallels of latitude, u_0 and u_1. Only those portions of the curves which lie in the shaded region correspond to real motion. All of these loci are dependent upon the values of u_0, a, and γ. However, their general character remains unchanged throughout large ranges in the values of these parameters, so that the history of gyroscopic motion (which is subject to the restrictions enumerated on p. 737) can be represented by means of eleven u, ψ_0-diagrams shown in Figs. I to XI, pp. 760–763. These diagrams are classified according to whether the quantities D, d, and u_0 are positive, negative, or zero, D and d being defined by the equations

$$D = \gamma^2 - a(1 + u_0), \qquad d = \gamma^2 - 2au_0. \tag{6}$$

Since each of the quantities D, d, and u_0 may be positive, negative, or zero, we should expect twenty-seven different cases. However, we have the inequality $d>D$, since $d-D=(1-u_0)>0$. This inequality eliminates twelve of the twenty-seven possibilities. Of the remaining fifteen, four are eliminated by the fact that $d>0$ when $u_0\leqq 0$.

Let us see how the quantities d and D are related to the motion of the gyroscope.

We shall show that the curve $u=u_1$ has a single maximum and a single minimum. If $D\geqq 0$, the maximum occurs when $u_1=+1$ (see Figs. I to VI).

* In this section it will be assumed that $-1<u_0<+1$. The case $u_0=\pm 1$ will be treated in §4.

If a gyroscope be given initial conditions such that $D \geqq 0$ and ψ_0 has the value for which u_1 is a maximum, then for this value of ψ_0, $u_1 = +1$ and the curve Γ traced by the gyroscope is bounded merely by the parallel of latitude u_0 and the north pole. If D is actually greater than zero, the point ξ, which traces Γ, passes periodically through the north pole. If, however, D is equal to zero, the point ξ approaches but never reaches the north pole. Hence, when $D=0$, t must become infinite as u approaches $+1$. That is, $+1$ must be a double root of $f(u)$. This corresponds in the u, ψ_0-diagram, to the fact that the maximum of u_1 coincides with the minimum of the curve u_2 when $D=0$ (see Figs. IV, V, VI).

If $D<0$, then the maximum of u_1 is less than $+1$. Hence, for such a value of D, it is impossible for ξ ever to pass through or approach asymptotically the north pole.

The minimum of u_1 is always -1. Therefore, no matter what other initial conditions may be assigned, it is always possible to choose the longitudinal velocity, ψ_0, so that ξ will pass through the south pole.

A gyroscope executes steady precession on a parallel of latitude, u_0, when and only when u_0 is a double root of $f(u)$. A double root of $f(u)$ corresponds in the u, ψ_0-diagram, to a point of intersection of the curve $u=u_1$ with the line $u=u_0$. But the curve u_1 has two, one, or zero intersections with the line u_0 according as d is positive, zero, or negative. Hence, there are two, one, or zero values of the longitudinal velocity ψ_0 for which the gyroscope can execute steady precession on the parallel of latitude, u_0, according as d is positive, zero, or negative.

The equation $\gamma = Cr/A$ shows that the quantities D and d depend upon the spin of the gyroscope. Thus, if the spin is large, D and d are positive. If the spin is small, then D is negative. Furthermore, if u_0 is positive, the spin can be taken so small that d is negative. Hence, by a choice of the spin, we can determine whether the axis of the gyroscope can pass through or approach asymptotically the north pole, and whether or not the gyroscope can execute steady precession on the parallel of latitude, u_0.

Let us see what types of motion are possible when we start a given gyroscope in such a manner that initially its axis passes through a point of a parallel of latitude u_0, the latitudinal velocity, θ, is zero, and the spin is so great that both d and D are positive. This situation corresponds to Figs. I, II, or III according as u_0 is positive, zero, or negative. Let us assume for definiteness that $u_0 > 0$.

We shall first take ψ_0 equal to zero. In the u, ψ_0-plane, the points corresponding to this motion constitute the segment of the u-axis included between the line $u=u_0$ and the curve $u=u_1$ (see Fig. I). The point $(0, u_0)$

corresponds to the initial point of the motion. It will be observed that the lines $u=u_0$ and $u=u_L$ intersect in this point. This corresponds to the fact that both the latitudinal and the longitudinal velocities are zero initially. Since these initial velocities are both zero, Γ must be a curve with cusps, the first cusp occurring at the initial point of the motion and the others occurring periodically thereafter. The point $(0, u_0)$ corresponds to all of these cusps.

The fact that the curves $u=u_i$ and $u=u_k$ pass through the point $(0, u_0)$ is without great significance to this motion; this point cannot correspond to a point of inflection of Γ since it corresponds to a cusp. However, the fact that the curves $u=u_i$ and $u=u_k$ do not cut the interior of the segment described above, shows that Γ does not have inflection points and that the geodesic curvature of Γ is monotone between points of contact with the parallels of latitude, u_0 and u_1.

Next let us take ψ_0 small but positive. Since the line $u=u_L$ does not enter the shaded region immediately to the right of the u-axis, it follows that for motions corresponding to such values of ψ_0, the longitudinal velocity never vanishes. Furthermore, since the longitudinal velocity is positive when u is equal to u_0, it is always positive. Thus Γ must be a wavy curve on which the point ξ moves eastward. This fact is indicated in the u, ψ_0-diagram by shading which slants to the right. The curves $u=u_i$ and $u=u_k$ enter the shaded region to the right of the u-axis and therefore a curve Γ corresponding to a small but positive value of ψ_0 has points of inflection and its geodesic curvature is not monotone between points of tangency to the parallels of latitude u_0 and u_1.

As we increase ψ_0, we reach a value corresponding to which the curve $u=u_k$ crosses the curve $u=u_1$ and leaves the shaded region. For this value of ψ_0 (and slightly greater values), the curve Γ is still wavy and still has points of inflection but the geodesic curvature of Γ is now monotone between points of tangency to the limiting parallels of latitude.

As ψ_0 is further increased, a value, $a/(2\gamma)$, is reached corresponding to which the curve $u=u_i$ crosses the line $u=u_0$ and leaves the shaded region. For this value of ψ_0, Γ no longer has inflection points; it is, however, still a wavy curve and the geodesic curvature is monotone between points of tangency to the limiting parallels of latitude. The same is true for slightly greater values of ψ_0.

Next we reach the point at which the curve $u=u_1$ intersects the straight line $u=u_0$. For this value of ψ_0, the limiting parallels of latitude, u_0 and u_1, coincide, Γ itself reduces to a parallel of latitude, u_0, and the gyroscope executes steady precession with positive longitudinal velocity.

If ψ_0 is still further increased, we reach a portion of the shaded region which does not contain the curve $u=u_i$ nor $u=u_k$ nor the straight line $u=u_L$. For such a value of ψ_0, the curve Γ is a wavy curve on which ξ moves eastward, it is without points of inflection and its geodesic curvature is monotone between points of tangency to the limiting parallels of latitude.

As we proceed further to the right, we reach a point at which the curve $u=u_i$ crosses the curve $u=u_1$ and reenters the shaded region. Hence we get again curves Γ with inflection points. As we proceed still further we reach a point at which the curve $u=u_k$ crosses the line $u=u_0$ and reenters the shaded region. Hence we have again curves Γ for which the geodesic curvature fails to be monotone between points of tangency to the limiting parallels of latitude.

When $\psi_0=a/\gamma$, the line $u=u_L$ and the curves $u=u_1$, $u=u_i$, and $u=u_k$ are concurrent. The curves $u=u_i$ and $u=u_k$ leave the shaded region and do not reenter at any point to the right of a/γ. Hence for $\psi_0\geqq a/\gamma$, the curves Γ are without points of inflection and their geodesic curvatures are monotone between points of tangency to their limiting parallels of latitude. The intersection of the line $u=u_L$ with the curve $u=u_1$ corresponds to the fact that the velocity, v, vanishes. Hence when $\psi_0=a/\gamma$, Γ is a curve with cusps.

The line $u=u_L$ enters the shaded region to the right of the line $\psi_0=a/\gamma$. Hence the longitudinal velocity vanishes between points of tangency of Γ to the limiting parallels of latitude. We shall prove that the longitudinal velocity changes sign whenever it vanishes, provided the value of u at which it vanishes lies in the open interval from u_0 to u_1. Assuming this to be true, it follows that the longitudinal velocity is negative at the point of tangency of Γ to the parallel of latitude u_1 since it is positive at the points of tangency of Γ to the parallel of latitude u_0. Thus Γ is a curve with loops. This type of motion is indicated by vertical shading.

If $\psi_0=\gamma/(1+u_0)$, then $u_1=+1$. Moreover, the curve $u=u_L$ does not pass through the point $(\gamma/(1+u_0), +1)$. Hence $+1$ is a root but not a double root of $f(u)$. It follows that for this value of ψ_0, Γ passes periodically through the north pole. It should be observed that, although $\dot{u}$ vanishes, at the north pole, $\dot{\theta}$ does not (when $D>0$). In fact, v must be different from zero at the north pole, since the contrary assumption leads to the conclusion that $+1$ is a double root of $f(u)$ when $\psi_0=\gamma/(1+u_0)$.

The line $u=u_L$ leaves the shaded region at the point $(\gamma/(1+u_0), +1)$ and does not enter at any point for which $\psi_0>\gamma/(1+u_0)$. It follows that, except when $u_0=u_1$, for such values of ψ_0, Γ is a wavy curve on which the point ξ moves eastward. But $u_0=u_1$ for only one value of ψ_0 greater than

$\gamma/(1+u_0)$. For this value of ψ_0, the gyroscope executes steady precession on the parallel of latitude, u_0.

The curves $u=u_i$ and $u=u_k$ do not enter the shaded region to the left of the u-axis. Hence a curve Γ which corresponds to a negative value of ψ_0 is without inflection points and the geodesic curvature of such a curve is monotone between points of tangency to its limiting parallels of latitude. The straight line $u=u_L$, on the other hand, enters the region to the left of the u-axis. Hence for a small negative value of ψ_0, the corresponding curve Γ is a curve with loops.

When $\psi_0=-\gamma/(1-u_0)$, $u_1=-1$. Thus for this value of ψ_0, the curve Γ passes through the south pole. The line $u=u_L$ leaves the shaded region at the point $(-\gamma/(1-u_0), -1)$ and does not reenter at any point for which $\psi_0<-\gamma/(1-u_0)$. It follows that all curves for which $\psi_0<-\gamma/(1-u_0)$, are wavy curves on which ξ moves westward.

Let $\{s_0s_1\}$ be an arc of Γ included between a point of tangency of Γ to the parallel of latitude u_0 and the next subsequent point of tangency to the parallel u_1. It will be recalled that if ψ_0 is such that either $\psi_0<0$ or $\psi_0>a/\gamma$, then on the corresponding arc $\{s_0s_1\}$, neither k nor k' vanishes at an interior point. It will be proved that such an arc lies entirely within its osculating circle at one extremity and entirely without its osculating circle at the other extremity. Within is understood to mean in that one, with the least area, of the two regions into which the given circle divides the spherical surface.

Figs. II to XI can be interpreted in a similar manner. Figs. I to XI are plotted for the following values of the parameters u_0, a, and γ:

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
u_0	.2	0	−.2	.3	0	−.3	.2	0	−.2	.2	.2
a	.8	.8	.8	1.3	1	.7	2	2	2	.1	.1
γ	1	1	1	1.3	1	.7	1	1	1	.2	.1

The general character of any one of these figures will remain unchanged if we make a variation in the values of u_0, a, γ, which does not alter the classification of this figure.

Let us investigate whether the data we have selected determines a reasonable physical situation. We may take as our gyroscope a homogeneous cone with altitude H, and radius of base R. For such a gyroscope we have the equations

$$a=\frac{5g}{R}\cdot\frac{\lambda}{1+\lambda^2},\qquad \gamma=\frac{2r}{1+\lambda^2}$$

where $\lambda=2H/R$. We shall choose the altitude to be two and a half times the

diameter of the base. Then $\lambda = 10$. In Case XI, $a = .1$, and hence for this case, the radius of the base of our gyroscope is about 4850 cm.

We do not, however, need to deal with such an enormous gyroscope. In fact, we can alter the quantities a and γ without changing the classification of the motion provided we keep u_0 and the ratio $\rho = a/\gamma^2$ fixed. Since both a and γ^2 have the dimensions T^{-2}, such an alteration may be brought about by a change in the unit of time. The effect of such a change, upon the u, ψ_0-diagram, is a uniform stretching in the direction of the ψ_0- axis.

We can now choose $\lambda = 10$, $a = 100$, and $\gamma = 10/\rho^{1/2}$. We obtain a conical gyroscope with altitude 24.25 cm. and diameter of base 9.7 cm. In Figs. I to XI, ρ varies from about .78 to 10. Hence for the corresponding motions the gyroscope makes from about 25 to 90 revolutions per second about its axis.

4. **Some exceptional cases.** In §3, it was assumed that $-1 < u_0 < +1$. It remains to consider the possibilities $u_0 = \pm 1$. In order to study these possibilities, we shall form the function*

$$\frac{f(u)}{u_0 - u} = (u_0 + u)[v_0^2 + a(u_0 - u)] - \gamma^2(u_0 - u) = g(u). \tag{7}$$

Since $g(u_0) = 2u_0v_0^2$ and $g(-u_0) = -2u_0\gamma^2$, it follows that $g(-1) \leqq 0$ and $g(+1) \geqq 0$ and, since $g(+\infty) = -\infty$, there is a root, u_1, of $g(u)$ in the interval $-1 \leqq u_1 \leqq +1$, and a root $u_2 \geqq +1$. Moreover, $u_0 = u_1$ or u_2 when and only when $v_0 = 0$, and $u_0 = -u_1$ or $-u_2$ when and only when $\gamma = 0$.

When $u_1 = u_0 = \pm 1$ or $u_2 = u_0 = +1$, the equation $\dot{\sigma} = \mu$ may be satisfied by setting $v \equiv 0$ and $u \equiv u_0$. For, when $u = \pm 1$, $\mu = 0$ and, when $v \equiv 0$ and r is constant, σ is a constant vector. In this case, the axis of the gyroscope remains vertical.

If $\gamma = 0$, then $r = 0$ and, if $u_0 = \pm 1$, then $\dot{\psi} \equiv 0$. Thus, if $v_0 \neq 0$, the gyroscope becomes a plane pendulum.† There are several cases to consider, namely, if $u_0 = -1$, then either u_1 or u_2 (or both) is equal to $+1$, and, if $u_0 = +1$, then $u_1 = -1$ and $u_2 > +1$.

First let us take $u_0 = -1$ and $u_2 = +1 > u_1$. In this case, the amplitude of the swing is $2\theta_1$ where $\cos \theta_1 = u_1$. Next let us take $u_0 = -1$ and $u_1 = u_2 = +1$. In this case, the angular velocity of the pendulum at the south pole is such that the pendulum approaches but never reaches the north pole.‡ Next

* In equation (7) it is assumed that $u_0 = \pm 1$. If $u_0 \neq \pm 1$, then the formula for $g(u) = f(u)/(u_0 - u)$ is given by equation (8).

† Cf. Love, *Theoretical Mechanics*, p. 129.

‡ Cf. Love, loc. cit., p. 131.

let us take $u_0=-1$ and $u_1=+1<u_2$. In this case, the angular velocity of the pendulum at the south pole is such that the pendulum passes through the north pole and swings completely around.* Finally, let us take $u_0=+1<u_2$ and $u_1=-1$. This case is the same as the case just discussed.

If $u_0=\pm 1$ and neither v_0 nor γ is zero, then $-1<u_1<+1<u_2$, and hence we can interchange the rôles of u_0 and u_1. That is, this case is equivalent to the one in which $-1<u_0<+1$.

5. **The bounding parallels of latitude.** In the u, ψ_0-diagram, the locus of points corresponding to points of tangency of Γ to the parallel of latitude u_0, is the straight line whose equation is $u=u_0$. Since u_1 is also a root of $f(u)$, the equation of the curve $u=u_1$ can be obtained by solving the equation $g(u)=0$ where $g(u)$ is defined by the equation

$$(8)\qquad \begin{aligned} g(u) &= f(u)/(u_0-u) \\ &= -au^2+[(1-u_0^2)\psi_0^2+\gamma^2]u+[(1-u_0^2)\psi_0^2u_0 \\ &\qquad +a-2(1-u_0^2)\gamma\psi_0-\gamma^2u_0]. \end{aligned}$$

Likewise, u_2 is a root of $g(u)$. Solving equation (8), we get

$$(9)\qquad u_1=\frac{(1-u_0^2)\psi_0+\gamma^2-(R(\psi_0))^{1/2}}{2a},\quad u_2=\frac{(1-u_0^2)\psi_0+\gamma^2+(R(\psi_0))^{1/2}}{2a},$$

where

$$\begin{aligned} R(\psi_0) &= (1-u_0^2)^2\psi_0^4+2(1-u_0^2)(\gamma^2+2au_0)\psi_0^2-8a\gamma(1-u_0^2)\psi_0 \\ &\quad +(\gamma^4-4a\gamma^2u_0+4a^2). \end{aligned}$$

In order to study the nature of the curve $u=u_1$, we shall compute its maxima and minima, treating, for convenience, u_1 as an implicit function defined by the equation $g(u)=0$ instead of making use of its explicit expression. Thus

$$\frac{du_\nu}{d\psi_0}=-\left.\frac{\dfrac{\partial g}{\partial\psi_0}}{\dfrac{\partial g}{\partial u}}\right|_{u=u_\nu}=\frac{-2(1-u_0^2)[(u_0+u_\nu)\psi_0-\gamma]}{(1-u_0^2)\psi_0^2-2au_\nu+\gamma^2}\qquad \text{where } \nu=1,2.$$

Setting the numerator equal to zero, we obtain the equation $u_\nu=\gamma/\psi_0-u_0$, and hence we obtain the following value for $g(u_\nu)$:

$$(10)\qquad g(\gamma/\psi_0-u_0)=\frac{-\gamma(1-u_0^2)}{\psi_0^2}\left(\psi_0-\frac{\gamma}{1+u_0}\right)\left(\psi_0+\frac{\gamma}{1-u_0}\right)\left(\psi_0-\frac{a}{\gamma}\right).$$

* Cf. Love, loc. cit., p. 131.

The roots of this last expression give the values of ψ_0 at the extrema of u_1 and u_2. The extrema are thus seen to occur at the points

$$(11)\quad P_1\colon \left(\frac{-\gamma}{1-u_0}, -1\right),\quad P_2\colon \left(\frac{\gamma}{1+u_0}, +1\right),\quad P_3\colon \left(\frac{a}{\gamma}, \frac{\gamma^2}{a}-u_0\right).$$

In order to compare the points P_2 and P_3, we shall write the coördinates of P_3 in the form

$$P_3\colon \left(\frac{\gamma}{1+u_0}-\frac{D}{\gamma(1+u_0)},\quad 1+\frac{D}{a}\right),$$

where D is defined by equation (6).

When $du_r/d\psi_0=0$,

$$\frac{d^2u_r}{d\psi_0^2} = -\frac{\partial^2 g}{\partial\psi_0^2}\Big/\frac{\partial g}{\partial u}\Big|_{u=u_r}.$$

Hence,

$$\text{at } P_1,\quad \frac{d^2u_r}{d\psi_0^2} = \frac{(1-u_0^2)(1-u_0)^2}{\gamma^2+a(1-u_0)} > 0\,;$$

$$\text{at } P_2,\quad \frac{d^2u_r}{d\psi_0^2} = \frac{-(1-u_0^2)(1+u_0)^2}{D},$$

$$\text{at } P_3,\quad \frac{d^2u_r}{d\psi_0^2} = \frac{2(1-u_0^2)\gamma^4}{a[\gamma^2+a(1-u_0)]D}.$$

P_1 is always a minimum but P_2 is a maximum and P_3 is a minimum if D is positive, and P_2 is a minimum and P_3 is a maximum if D is negative. It should be observed that $\partial g/\partial u$ never vanishes at P_1 and that it vanishes at P_2 and P_3 only when $D=0$. But, when $D=0$,

$$(12)\quad u_1 = \begin{cases} 1+\left(\psi_0-\dfrac{a}{\gamma}\right)\lambda_1(\psi_0) & \text{if } \psi_0 < a/\gamma, \\ 1+\left(\psi_0-\dfrac{a}{\gamma}\right)\lambda_2(\psi_0) & \text{if } \psi_0 > a/\gamma, \end{cases}$$

$$u_2 = \begin{cases} 1+\left(\psi_0-\dfrac{a}{\gamma}\right)\lambda_2(\psi_0) & \text{if } \psi_0 < a/\gamma, \\ 1+\left(\psi_0-\dfrac{a}{\gamma}\right)\lambda_1(\psi_0) & \text{if } \psi_0 > a/\gamma, \end{cases}$$

where

$$(13)\quad \lambda_1(\psi_0) = \frac{1}{a}\left[(1-u_0)\gamma + \frac{1}{2}(1-u_0^2)(\psi_0-a/\gamma) + r(\psi_0)\right],$$

$$\lambda_2(\psi_0) = \frac{1}{a}\left[(1-u_0)\gamma + \frac{1}{2}(1-u_0^2)(\psi_0-a/\gamma) - r(\psi_0)\right],$$

and

$$(14) \qquad r(\psi_0) = \left(\left[\gamma(1-u_0) + \frac{1}{2}(1-u_0^2)(\psi_0 - a/\gamma)\right]^2 + (1-u_0^2)\gamma^2\right)^{1/2}.$$

Equations (12) may be derived from equation (9) by repeated use of the condition $D=0$. When this condition is fulfilled, the points P_2 and P_3 coincide, and, if we rewrite this condition in the form $a/\gamma=\gamma/(1+u_0)$, equation (12) shows that this point lies on both u_1 and u_2. This point is a maximum of u_1 and a minimum of u_2 (as we have defined these functions).

We have proved that u_1 and u_2 are real and that $-1 \leqq u_1 \leqq +1 \leqq u_2$ for all values of ψ_0. Furthermore, when $D>0$, u_1 has a maximum at P_2, and P_3 lies above and to the left of P_2; when $D<0$, u_1 has a maximum at P_3 and P_3 lies below and to the right of P_2; when $D=0$, the points P_2 and P_3 coincide at a maximum of u_1 and a minimum of u_2.

The point P_1 corresponds to a passage of the curve Γ through the south pole. When $D>0$, P_2 corresponds to a passage of Γ through the north pole. When $D=0$, u_1 and u_2 coincide at P_2 which corresponds to an asymptotic (with respect to t) approach to the north pole.* If $D<0$, P_2 does not correspond to real motion.

Another property of the function $u_1(\psi_0)$ is that it approaches $-u_0$ from above as ψ_0 becomes positively infinite and from below as ψ_0 becomes negatively infinite. This is proved as follows. Let $x=1/\psi_0$, $h(u, x)=x^2g(u)$. If $x\neq 0$, the roots of g and h are the same. But $\partial h(-u_0, 0)/\partial u = 1-u_0^2 > 0$. Hence $h(u, x)=0$ defines a function $u=u(x)$ which is continuous at the point $(0, -u_0)$. Therefore

$$\lim_{x\to 0} u(x) = \lim_{\psi_0 \to \pm\infty} u_1 = -u_0,$$

and, since the derivative is positive at $x=0$, u_1 takes on values slightly greater than or slightly less than $-u_0$ according as x takes on values slightly greater than or slightly less than zero or according as ψ_0 is large and positive or large and negative.

This situation is in accordance with what should be expected from dynamical considerations. For, as ψ_0 becomes large, v also becomes large and the gyroscopic force and the force of gravity become relatively unimportant. Hence the path of the gyroscope becomes approximately a great circle. But, if a great circle is drawn tangent to a parallel of latitude $u=u_0$, then it must also be tangent to the parallel of latitude $u=-u_0$, the two points of tangency being at opposite ends of a diameter of the great circle. Thus we would

* Cf. Routh, *Advanced Rigid Dynamics*, 1905, pp. 136, 137.

expect that, as ψ_0 becomes infinite, the geodesic curvature, k, must approach zero, and the difference in longitude, $\Delta\psi$, between a point of tangency of Γ with u_0 and the next subsequent point of tangency of Γ with u_1 must approach π.

An expression for k may be obtained by solving equation (1b) and combining equations (2) and (4). Thus*

$$k = \frac{\gamma}{v^3}\left[(1-u_0^2)\psi_0^2 - \frac{a}{2\gamma}(1-u_0^2)\psi_0 + \frac{a}{2}(u_0-u)\right] \tag{15}$$

and, since $v^2=(1-u_0^2)\psi_0^2+a(u_0-u)$, k must approach zero as ψ_0 becomes positively or negatively infinite.

The difference in longitude, $\Delta\psi$, may be obtained from equations (4) and (5). Thus

$$\Delta\psi = \int_{u_0}^{u_1} \frac{(1-u_0^2)\psi_0 + \gamma(u_0-u)}{(1-u^2)(f(u))^{1/2}}du.$$

This expression may be written in the form

$$\Delta\psi = \frac{(1-u_0^2)+\gamma(u_0+\bar{u})/\psi_0}{\left((1-u_0^2)+\epsilon\left(\frac{1}{\psi_0},\bar{u}\right)\right)^{1/2}}\int_{u_0}^{u_1}\frac{du}{(1-u^2)(u_0^2-u^2)^{1/2}},$$

where $u_0<\bar{u}<u_1$ and

$$\epsilon\left(\frac{1}{\psi_0},\bar{u}\right) = \frac{(1/\psi_0)[a(1-u^2)-\gamma^2(u_0-\bar{u})]-2\gamma(u_0-u)}{\psi_0(u_0+\bar{u})}.$$

Since

$$\lim_{\psi_0\to\pm\infty}\frac{(1-u_0^2)+\gamma(u_0+\bar{u})/\psi_0}{((1-u_0^2)+\epsilon(1/\psi_0,\bar{u}))^{1/2}} = (1-u_0)^2 \quad\text{and}\quad \lim_{\psi_0\to\pm\infty} u_1 = -u_0 > 0,$$

and since

$$\int_{+u_0}^{-u_0}\frac{du}{(1-u^2)(u_0^2-u^2)^{1/2}} = \pi/(1-u_0^2)^{1/2} \quad\text{when}\quad u_0<0,$$

it follows that

$$\lim_{\psi_0\to\pm\infty}\Delta\psi = \pi.$$

We proceed similarly when $u_0>0$.

The curve $u=u_1$ is further characterized by its intersections with the line $u=u_0$. These points of intersection are of special dynamical interest

* Cf. Osgood, loc. cit., p. 258, equation IV.

since the condition $u_1 = u_0$ is necessary and sufficient for steady precession.* The values of ψ_0 at which the curve $u = u_1$ crosses the straight line $u = u_0$ are given by solving the equation $g(u_0) = 0$ for ψ_0. Thus

$$\psi_0 = \frac{\gamma \pm d^{1/2}}{2u_0} \tag{16}$$

where d is defined by equation (7) and where $u_0 \neq 0$. These values of ψ_0 are real and distinct, conjugate imaginary or coincident according as d is positive, negative, or zero. If $u_0 = 0$ the solution of $g(u_0) = 0$ is $\psi_0 = a/(2\gamma)$.

6. **Paths for which the longitudinal velocity vanishes.** If the path, Γ, of the gyroscope is a curve with loops, then the longitudinal velocity, ψ_0, when $u = u_0$, must be opposite in sign to the longitudinal velocity, ψ_1, when when $u = u_1$, and hence ψ must vanish for some value of u between u_0 and u_1. If, at certain points, Γ has cusps, then, at such points, v must vanish, since by equation (16), k can be discontinuous only when $v = 0$. But, when v vanishes, $\dot{u}$ and ψ must both be equal to zero. Hence the character of Γ is related to the vanishing or non-vanishing of ψ.

If $\psi = 0$, we may solve equation (4) for u and get

$$u = u_L(\psi_0) = u_0 + (1 - u_0^2)\frac{\psi_0}{\gamma}. \tag{17}$$

The equation $u = u_L$ represents a straight line intersecting the line $u = u_0$ in the point P_0: $(0, u_0)$. The intersections of $u = u_L$ with the curve $u = u_1$ are given by solving $g(u_L) = 0$ for ψ_0. Substituting $u = u_L$ and factoring, we get

$$g(u_L) = \frac{1}{\gamma}(1 - u_0^2)^2\left[\psi_0 - \frac{\gamma}{1 + u_0}\right]\left[\psi_0 + \frac{\gamma}{1 - u_0}\right]\left[\psi_0 - \frac{a}{\gamma}\right].$$

Hence the points of intersection of $u_L(\psi_0)$ with $u_1(\psi_0)$ are P_1: $(-\gamma/(1-u_0), -1)$, P_2: $(\gamma/(1+u_0), +1)$ and P_4: $(a/\gamma, u_0+(1-u_0^2)a/\gamma^2)$. We may write the coördinates of P_4 in the form P_4: $(\gamma/(1+u_0)-D/[\gamma(1+u_0)], 1-D(1-u_0)/\gamma^2)$. Hence if $D \neq 0$, P_4 lies on the curve $u_1(\psi_0)$ when P_3 lies on the curve $u_2(\psi_0)$ and P_4 lies on the curve $u_2(\psi_0)$ when P_3 lies on the curve $u_1(\psi_0)$, and, if $D = 0$, P_2, P_3, and P_4 coincide. The line $u = u_L$ is shown in Figs. I to XI.

When $0 < \psi_0 < a/\gamma$, the line $u_L(\psi_0)$ lies outside of the region between u_0 and u_1, since u_L lies outside of this region for ψ_0 small and positive, and does not cross u_0 or u_1 between P_0 and P_4. The line u_L lies inside the region between u_0 and u_1 between the points P_0 and P_1 and, when $D < 0$, between P_2

* Cf. p. 738.

and P_4. All other parts of u_L lie outside of this region. Hence the values of ψ_0 corresponding to the points of u_L which lie in this region, are given by the inequalities*

$$\frac{-\gamma}{1-u_0} \leqq \psi_0 \leqq 0 \quad \text{or} \quad \frac{a}{\gamma} \leqq \psi_0 \leqq \frac{\gamma}{1+u_0}.$$

If ψ_0 is such that Γ is a curve with loops, then we have seen that, for this value of ψ_0, the point $(\psi_0, u_L(\psi_0))$ must lie in the open region included between u_0 and $u_1(\psi_0)$. The converse is also true; namely, if $(\psi_0, u_L(\psi_0))$ lies in the open region between u_0 and u_1, then ψ_0 and ψ_1 have opposite signs and hence Γ is a curve with loops. For, by equation (4),

$$\ddot{\psi}_0 = \frac{-\gamma\dot{u}}{1-u^2} \quad \text{when } \psi = 0,$$

and thus ψ and $\ddot{\psi}$ cannot vanish simultaneously when u lies in the open interval between u_0 and u_1. Therefore ψ changes sign whenever it vanishes. But ψ vanishes for, at most, one value of u between u_0 and u_1, since $u=u_L(\psi_0)$ is a single-valued function. Hence ψ_0 and ψ_1 have opposite signs when and only when $(\psi_0, u_L(\psi_0)$ lies between $u=u_0$ and $u=u_1$, and consequently when and only when $-\gamma/(1-u_0)<\psi_0<0$ or $a/\gamma<\psi_0<\gamma/(1+u_0)$. It follows that ψ_1 is negative when $\psi_0 < -\gamma/(1-u_0)$ and positive when $-\gamma/(1-u_0)<\psi_0\leqq 0$. If $D>0$, ψ_1 is positive when $0\leqq\psi_0<a/\gamma$, negative when $a/\gamma<\psi_0<\gamma/(1+u_0)$, and positive when $\gamma/(1+u_0)<\psi_0$. If $D=0$, ψ_1 is positive when $0\leqq\psi_0<\gamma/(1+u_0)$ and when $\gamma/(1+u_0)<\psi_0$. If $D<0$, ψ_1 is positive when $0\leqq\psi_0$. Consequently Γ is a wavy curve for which the drift of the motion is westward when $\psi_0<-\gamma/(1-u_0)$, and Γ is a wavy curve for which the drift of the motion is eastward when $D>0$ and $0<\psi_0<a/\gamma$ or $\gamma/(1+u_0)<\psi_0$, when $D=0$ and $0<\psi_0<\gamma/(1+u_0)$ or $\gamma/(1+u_0)<\psi_0$, or when $D<0$ and $\psi_0>0$. Γ is a curve with loops when $-\gamma/(1-u_0)<\psi_0<0$ or $a/\gamma<\psi_0<\gamma/(1+u_0)$.

7. **Paths having points of inflection.** The purpose of this paragraph is to determine what initial conditions are necessary in order that Γ may have points of inflection. At such points, the geodesic curvature, k, must vanish. But, when k is zero, we may solve equation (15) for u and get

$$u = u_0 + (1-u_0^2)\frac{2\psi_0}{a}(\psi_0 - a/(2\gamma)) = u_i(\psi_0). \tag{18}$$

The curve $u=u_i$ is a parabola with its concave side upward. Only those portions of this parabola which lie in the region between $u=u_0$ and $u=u_1$ correspond to real motion. The intersections of the curve u_i with the bound-

* The inequality $a/\gamma\leqq\gamma/(1-u_0)$ implies $D\geqq 0$.

ary of this region are obtained by solving the equation $g(u_i)=0$ for ψ_0 where

$$g(u_i) = \frac{-2}{a}(1-u_0^2)(\psi_0 - a/\gamma)h(\psi_0),$$

(19)

$$h(\psi_0) = (1-u_0^2)\psi_0^3 - \frac{a}{2\gamma}(1-u_0^2)\psi_0^2 - (\gamma^2 - au_0)\psi_0 + \frac{a\gamma}{2}.$$

We shall study the roots of $h(\psi_0)$. The signs of the coefficients of the powers of ψ_0 in $h(\psi_0)$ are given in the following scheme:

$$\begin{aligned} &+\ -\ -\ + \text{ when } \gamma^2 - au_0 > 0,\\ &+\ -\ + \text{ when } \gamma^2 - au_0 = 0,\\ &+\ -\ +\ + \text{ when } \gamma^2 - au_0 < 0. \end{aligned}$$

In all three cases the number of sign changes is two and therefore (by Descartes' rule of signs) the number of positive roots is at most two. If we replace ψ_0 by $-\psi_0$, the signs of the coefficients are given by the following scheme:

$$\begin{aligned} &-\ -\ +\ + \text{ when } \gamma^2 - au_0 > 0,\\ &-\ -\ + \text{ when } \gamma^2 - au_0 = 0,\\ &-\ -\ -\ + \text{ when } \gamma^2 - au_0 < 0. \end{aligned}$$

Thus in all cases, the number of negative roots is at most one. Since $h(0)=\frac{1}{2}a\gamma>0$ and $h(-\infty)=-\infty$, $h(\psi_0)$ must have precisely one negative root and therefore the other two roots must be either both positive or else conjugate imaginary. If the roots are real, the question which concerns us is whether these roots correspond to intersections of u_i with u_1 or with u_2. In order to investigate this question we form the function

$$\phi(\psi_0) = u_i(\psi_0) - u_1(\psi_0). \tag{20}$$

We shall show that $\phi(0)=u_0-u_1(0)$ is positive. The condition $\phi(0)>0$ is equivalent to the statement that *if the longitudinal velocity on a limiting circle vanishes, then that circle is the upper limiting circle.* Let us denote by u_0 the limiting circle upon which the longitudinal velocity vanishes. Then $v^2=v_0^2+a(u_0-u)$. But the latitudinal velocity also vanishes when $u=u_0$ and therefore $v_0=0$. Thus, if u takes on greater values than u_0, v becomes imaginary, which is impossible. Therefore u_0 must be the upper limiting circle and $\phi(0)$ must be positive.*

Since $u_1(+\infty)=-u_0$, it follows that $\phi(+\infty)=+\infty$. But, since ϕ is everywhere continuous, if it has one positive root, it must have a second.

* This fact may also be proved directly from the explicit formula for $u_1(0)$.

The same holds true for the negative roots of ϕ. But the roots of ϕ are also roots of $g(u_i)$ and since

$$\frac{d}{d\psi_0}g[u_i(\psi_0)] = \frac{\partial g}{\partial u}\frac{du_i}{d\psi_0} + \frac{\partial g}{\partial \psi_0} = \frac{\partial g}{\partial u}\left(\frac{du_i}{d\psi_0} - \frac{du_i}{d\psi_0}\right) = \frac{\partial g}{\partial u}\frac{d\phi}{d\psi_0} \text{ when } u_i = u_1 = u,$$

the double roots of ϕ are also the double roots of $g(u_i)$. Therefore ϕ cannot have a negative root. Hence the negative root of $h(\psi_0)$ corresponds to an intersection of $u_i(\psi_0)$ with $u_2(\psi_0)$.

As we proceed to the left, starting at $\psi_0=0$, the curve u_i does not enter the region between u_0 and u_1, since initially u_0 is greater than u_1 and, as ψ_0 decreases, u_i increases from the value u_0. Also, since the curve u_i does not cross the line u_0 or the curve u_1 at a negative value of ψ_0, it lies completely without the region bounded by u_0 and u_1 to the left of the ψ_0-axis. But, since our choice of u_0 was arbitrary, we have proved that Γ *cannot have a point of inflection if the longitudinal velocity on either limiting circle is negative.*

Formula (19) shows that

$$\psi_0 = a/\gamma = \frac{\gamma}{1+u_0} - \frac{D}{\gamma(1+u_0)}$$

is a root of $g(u_i)$; the corresponding value of u is

$$u_0 + (1-u_0^2)a/\gamma^2 = 1 - \frac{(1-u_0)D}{\gamma^2}.$$

Hence the point

$$P_4\colon \left(\frac{\gamma}{1+u_0} - \frac{D}{\gamma(1+u_0)},\ 1 - \frac{(1-u_0)D}{\gamma^2}\right)$$

lies on the curve u_2, u_1, or both, according as D is less than, greater than, or equal to zero. Thus, if $D \leqq 0$, then $u_i(a/\gamma) \geqq +1$, and since u_i is an increasing function to the right of the line $\psi_0 = a/\gamma$, the curve u_i lies completely withou the region bounded by u_1 and u_0 to the right of a/γ.

If $D>0$, ψ_1 is negative for ψ_0 slightly greater than a/γ and thus the curve u_i cannot enter the region between u_0 and u_1 at a/γ as we proceed to the right, since Γ cannot have a point of inflection if the longitudinal velocity at $u=u_1$ is negative. But, since u_i and u_1 have definite slopes at a/γ, the curve u_i must enter the region between u_0 and u_1 as we proceed to the left at a/γ, or else u_i and u_1 must have the same slope at this point. The latter possibility is excluded since it implies a double root of ϕ at a/γ which is contrary to the fact that

$$\begin{aligned} h(a/\gamma) &= (1-u_0^2)a^3/(2\gamma^3) - (\gamma^2 - au_0)a/\gamma + \tfrac{1}{2}a\gamma \\ &= -D[\gamma^2 + a(1-u_0)]a/(2\gamma^3) \neq 0 \quad \text{when } D > 0. \end{aligned}$$

Thus, when D is positive, ϕ must have a root between 0 and a/γ. Furthermore, ϕ must be positive for values of ψ_0 slightly greater than a/γ and therefore, if ϕ has one root greater than a/γ, it must have two such roots (since $\phi(+\infty)=+\infty$), and hence must have four positive roots. But ϕ can have at most three positive roots (the root a/γ and the two positive roots of h). Therefore ϕ cannot have a root greater than a/γ even when D is positive. We have thus proved that *a necessary condition that* Γ *have a point of inflection is that* ψ_0 (*and incidentally* ψ_1) *lie in the closed interval from* 0 *to* a/γ.

The parabola u_i goes below the line u_0 at $\psi_0=0$ and comes above the line again at $a/(2\gamma)$. The curve u_1 may cut this parabola in one, two, or no points in the interval from 0 to $a/(2\gamma)$. The first possibility always occurs when D is positive. The second possibility may be seen by observing that the minimum of u_i is at $a/(4\gamma)$ and $u_i(a/(4\gamma))=u_0-(1-u_0^2)a/(8\gamma^2)$. This may be made less than -1 by choosing γ sufficiently small, and if such a choice of γ is made, u_1 must certainly cross u_i in the interval from 0 to $a/(2\gamma)$. The third possibility may be shown by a numerical example. Thus, let $u_0=\frac{1}{2}$, $a=4$, and $\gamma=1$. Then $h(\psi_0)=\frac{3}{4}\psi_0^3-\frac{3}{2}\psi_0^2+\psi_0+2$ and this expression has no positive roots.

We have proved that when u lies in the closed interval from u_0 to u_1, k can vanish only when $0\leqq\psi_0\leqq a/\gamma$. Hence for all other values of ψ_0, k has the same sign as k_0 (where k_0 is the value of k when $u=u_0$). But $k_0=(\gamma/v_0^3)\cdot(1-u_0^2\psi)_0(\psi_0-a/(2\gamma))$ and this expression is positive whenever $\psi_0<0$ or $\psi_0>a/\gamma$. *Hence* k *is always positive except when* $0\leqq\psi_0\leqq a/\gamma$. *Therefore, only wavy curves for which the drift of the motion is eastward can have points of inflection.*

8. **Paths having monotone geodesic curvatures.** If the derivative of the geodesic curvature with respect to the arc length of the path does not vanish in any given interval, then the geodesic curvature is monotone in that interval. An expression for k' may be obtained by differentiating equation (15) and combining terms by means of equation (3). Thus*

$$k' = \frac{+\gamma a\dot{u}}{2v^6}L(u,\psi_0),$$

(21)

$$L(u,\psi_0) = 2(1-u_0^2)\psi_0^2 - \frac{3a}{2\gamma}(1-u_0^2)\psi_0 + \frac{a}{2}(u_0-u).$$

Setting $L(u,\psi_0)$ equal to zero, we obtain the equation

$$u = u_0 + \frac{4}{a}(1-u^2)\psi_0\left(\psi_0 - \frac{3a}{4\gamma}\right) = u_k(\psi_0).$$

* Cf. Kellogg, loc. cit., p. 519.

Hence k' vanishes when and only when $u=u_k(\psi_0)$, $u=u_0$, or $u=u_1(\psi_0)$.

The curve $u_k(\psi_0)$ is a parabola which intersects the parabola $u_i(\psi_0)$ in the points P_0: $(0,\ u_0)$ and P_4: $(a/\gamma,\ u_0+a(1-u_0^2)/\gamma^2)$, and only at these points. But P_0 lies on the line $u=u_0$ and P_4 lies on the curve $u=u_1$ or $u=u_2$. Hence k and k' vanish simultaneously (at points corresponding to real motion) only at the intersections of u_i with u_0 and u_1. These points of intersection cannot correspond to points of inflection of the path, since Γ is symmetric in any meridian circle through a point of contact with the parallel of latitude u_0 or u_1 and hence k cannot change sign at such a point. Therefore, in the u, ψ_0-plane, the points of the parabola $u=u_i$ which lie in the open region included between u_0 and u_1, and only these points, correspond to points of inflection of the path Γ.

We have proved that the parabola $u=u_i(\psi_0)$ lies inside the shaded region* between u_0 and u_1 only when $0\leqq\psi_0\leqq a/\gamma$. The same is true for the parabola u_k. This follows from the fact that the portion of u_k for which $\psi_0<0$ lies in the region bounded on the left by the parabola u_i and on the right by the ψ_0-axis. This region, we have seen, does not overlap the region between u_0 and u_1. Likewise, the portion of u_k for which $a/\gamma<\psi_0$, lies in the region bounded on the left by the line $\psi_0=a/\gamma$ and on the right by the parabola u_i. This region does not overlap the region between u_0 and u_1. Consequently, the path Γ has monotone geodesic curvature between a point of tangency with u_0 and the first subsequent point of tangency with u_1 (or vice versa), whenever $\psi_0<0$ or $a/\gamma<\psi_0$. The same is true whenever $\psi_1<0$ or $a/\gamma<\psi_1$.

Since $L(u,\ \psi_0)$ can vanish only when $0\leqq\psi_0\leqq a/\gamma$, and since $L(u_0,\ \psi_0)=2(1-u_0^2)\psi_0(\psi_0-3a/(4\gamma))$ is positive whenever $\psi_0<0$ or $\psi_0>a/\gamma$, it follows that sgn $k'=$sgn $\dot{u}$ whenever $\psi_0<0$ or $\psi_0>a/\gamma$.

O. D. Kellogg has proved† that if the arc derivative K' of the curvature‡, K, of an arc $\{s_1s_2\}$ (where $s_1<s_2$) of a spherical curve is always positive, then the arc $\{s_1s_2\}$ lies entirely within the osculating circle at s_1, and entirely without the osculating circle at s_2. The curvature, K, is related to the geodesic curvature, k, by means of the equation

$$1+k^2=K^2.$$

Hence $kk'=KK'$. But we have proved that, whenever $\psi_0<0$ or $\psi_0>a/\gamma$, $k>0$ and sgn $k'=$sgn $\dot{u}$. Therefore, when $\psi_0<0$ or $\psi_0>a/\gamma$, an arc of Γ

* Cf. Figs. I to XI.

† Cf. O. D. Kellogg, loc. cit., pp. 509, 521.

‡ The arc $\{S_1S_2\}$ is regarded as an arc of a space curve and K is the curvature at a given point of this arc.

consisting of a half wave lies entirely within the osculating circle at its point of tangency with u_0, and entirely without the osculating circle at its point of tangency with u_1, or entirely within the osculating circle at u_1 and entirely without the osculating circle at u_0, according as u_1 is north or south of u_0.

9. **The spherical pendulum.** When $\gamma=0$, the gyroscope becomes a spherical pendulum.* Thus we can discuss the behavior of the spherical pendulum if we know the nature of the curves u_1, u_L, u_i, and u_k.†

As γ approaches zero, the point P_1: $(-\gamma/(1-u_0), -1)$ approaches the point $(0, -1)$, the point P_2: $(\gamma/(1+u_0), +1)$ approaches the point $(0, +1)$, and the point $P_3(a/\gamma, \gamma^2/(a-u_0))$ goes out to $+\infty$. Hence the line u_L turns into coincidence with the u-axis and the curves u_i and u_k degenerate into the u-axis. Thus we have again the familiar facts that the paths of the spherical pendulum are always wavy curves and are without inflection points.‡ We have in addition the fact that the curvature of a path of a spherical pendulum is always monotone between points of tangency to the bounding parallels of latitude.

The curve u_i crosses the straight line u_0 at the points $\psi_0=\pm(-a/(2u_0))^{1/2}$. These points are real when and only when $u_0<0$. For these values of ψ_0, the phenomenon of the conical pendulum occurs. When $\psi_0=0$, we have a plane pendulum.

The curve u_i approaches the straight line $-u_0$ when ψ_0 becomes positively or negatively infinite. Thus, as in the case of the gyroscope, the force of gravity is relatively unimportant when ψ_0 is large.

10. **Curves near the north or south pole.** Let us consider a curve passing through the north pole. For this curve the longitudinal velocity at u_0 is $\psi_0=\gamma/(1+u_0)$. Hence equation (4) reduces to

$$\psi = \frac{\gamma}{1+u}.$$

Thus the change in longitude, $\Delta\psi$, between u_0 and the north pole is given by the integral

$$\Delta\psi = \int_{u_0}^{1} \frac{\gamma du}{(1+u)(a(u-u_0)(1-u)(u_2-u))^{1/2}}.$$

This integral may be written in the form

$$\Delta\psi = \frac{\gamma}{(a(u_2-u_0))^{1/2}} \int_{u_0}^{1} \frac{du}{(1+u)((u-u_0)(u_2-u))^{1/2}}$$

* Cf. Appell, loc. cit., Tome I, pp. 513–524.

† These curves are shown in Fig. XII.

‡ Cf. the photographs of the paths of spherical pendulums (Webster, *Dynamics*, pp. 50, 51).

where $u_0<\bar{u}<1$. Therefore

$$\Delta\psi = \frac{\pi\gamma}{(2a(1+u_0)(u_2-\bar{u}))^{1/2}}.$$

Next let us find an expression for u_2. This is easily done since, when $\psi_0=\gamma/(1+u_0)$, $f(u)=(1-u)\,(u_0-u)\,[a(1+u)-2\gamma^2/(1+u_0)]$. Thus we obtain the expression

$$u_2 = -1+\frac{2\gamma^2}{a(1+u_0)}.$$

Substituting this value in the formula for $\Delta\psi$, we obtain the equation

$$\Delta\psi = \frac{\pi}{2}[1-a(1+u_0)(1+\bar{u})/(2\gamma^2)]^{-1/2}.$$

Furthermore if we set $\gamma^2=2a$, then as u_0 approaches 1, $\Delta\psi$ becomes infinite. Hence, by a proper choice of a, γ and u_0, we can make $\Delta\psi$ as large as we please.

If we set $\psi_0=\gamma/(1+u_0)+h$, then, corresponding to any value of h, we obtain a path $\Gamma(h)$. Let us denote by $\Delta\psi(h)$ the change in longitude along $\Gamma(h)$ between u_0 and u_1. As h approaches zero either through positive or through negative values, the circle u_1 approaches the north pole and the curve $\Gamma(h)$ approaches the limiting curve $\Gamma(0)$. If $|h|$ is small and $h<0$, $\Gamma(h)$ is a curve with loops, and if $h>0$, $\Gamma(h)$ is a wavy curve. This situation is shown in Fig. XIII.

If $h>0$, then the change in longitude along the curve $\Gamma(h)$ from P_0 to P_1 is $2\Delta\psi(h)$. It is easily seen geometrically that $2\Delta\psi(h)-2\Delta\psi(0)$ differs from π by an amount which is numerically equal to the difference in longitude between P_1 and P_2. But, as h approaches zero, P_1 approaches P_2. Therefore,*

$$\lim_{h\to 0^+}\Delta\psi(h) = \Delta\psi(0)+\frac{\pi}{2}.$$

In a like manner, it may be proved that*

$$\lim_{h\to 0^-}\Delta\psi(h) = \Delta\psi(0)-\frac{\pi}{2}.$$

* These equations can also be obtained analytically. For we have the equation

$$\Delta\psi(h) = \int_{u_0}^{u_1}\frac{\gamma du}{(1+u)(a(u-u_0)(u_1-u)(u_2-u))^{1/2}}+\int_{u_0}^{u_1}\frac{(1-u_0^2)h du}{(1-u^2)(a(u-u_0)(u_1-u)(u_2-u))^{1/2}}.$$

As h approaches zero, u_1 approaches $+1$. Hence

$$\lim_{h\to 0}\int_{u_0}^{u_1}\frac{\gamma du}{(1+u)(a(u-u_0)(u_1-u)(u_2-u))^{1/2}} = \int_{u_0}^{1}\frac{\gamma du}{(1+u)(a(u-u_0)(1-u)(u_2-u))^{1/2}} = \Delta\Psi(0).$$

It remains to prove that

But we have proved that $\Delta\psi(0)$ may be made as large as we please and hence by choosing $|h|$ sufficiently small, $\Delta\psi(h)$ may be made as large as we please.*

$$\lim_{h\to 0}\int_{u_0}^{u_1}\frac{(1-u_0^2)hdu}{(1-u^2)(a(u-u_0)(u_1-u)(u_2-u))^{1/2}}=\pm\frac{\pi}{2}.$$

By means of the substitution $u=1-(1-u_1)x$, we obtain the equation

$$\int_{u_0}^{u_1}\frac{(1-u_0^2)hdu}{(1-u^2)(a(u-u_0)(u_1-u)(u_2-u))^{1/2}}$$
$$=\frac{h}{(1-u_1)^{1/2}}\int_1^{(1-u_0)/(1-u_1)}\frac{(1-u_0^2)dx}{x[2-(1-u_1)x](a[(1-u_0)-(1-u_1)x][(u_2-1)+(1-u_1)x][x-1])^{1/2}}.$$

Furthermore

$$\lim_{h\to 0}\int_1^{(1-u_0)/(1-u_1)}\frac{(1-u_0^2)dx}{x[2-(1-u_1)x](a[(1-u_0)-(1-u_1)x][(u_2-1)+(1-u_1)x][x-1])^{1/2}}$$
$$=\int_1^{\infty}\frac{(1-u_0^2)dx}{2x[(x-1)a(1-u_0)(u_2-1)]^{1/2}}=\frac{\pi}{2}\frac{1-u_0^2}{[a(1-u_0)(u_2-1)]^{1/2}}.$$

Finally we have to evaluate $\lim_{h\to\infty}[h/(1-u_1)^{1/2}]$. In order to obtain a value for this limit, we shall expand the function $u_1=u_1(\gamma/(1+u_0)+h)$ about the point $h=0$. But when $h=0$ (cf. p. 747)

$$\frac{du_1}{d\psi_0}=0 \text{ and } \frac{d^2u_1}{d\psi_0^2}=-\frac{(1-u_0^2)(1+u_0)^2}{D}.$$

Hence

$$u_1=1-\frac{(1-u_0^2)(1+u_0)^2}{D}h^2+O_3(h)$$

where $O_3(h)$ is an infinitesimal of at least the third order in h. Therefore

$$\lim_{h\to 0+}\frac{h}{(1-u_1)^{1/2}}=-\lim_{h\to 0-}\frac{h}{(1-u_1)^{1/2}}=\frac{(2D)^{1/2}}{(1+u_0)(1-u_0^2)^{1/2}}.$$

It follows that

$$\lim_{h\to 0+}\int_{u_0}^{u_1}\frac{(1-u_0^2)hdu}{(1-u_0^2)[a(u-u_0)(u_1-u)(u_2-u)]^{1/2}}=-\lim_{h\to 0-}\int_{u_0}^{u_1}\frac{(1-u_0^2)hdu}{(1-u_0^2)[a(u-u_0)(u_1-u)(u_2-u)]^{1/2}}$$
$$=\frac{\pi}{2}\left(\frac{2D}{a(1+u_0)(u_2-1)}\right)^{1/2}.$$

But $u_2=-1+2\gamma^2/[a(1+u_0)]$ and $D=\gamma^2-a(1+u_0)$ and thus

$$\left(\frac{2D}{a(1+u_0)(u_2-1)}\right)^{1/2}=1.$$

Consequently

$$\lim_{h\to 0+}\Delta\psi(h)=\Delta\psi(0)+\pi/2$$

and

$$\lim_{h\to 0-}\Delta\psi(h)=\Delta\psi(0)-\pi/2.$$

* In the case of the spherical pendulum, G. H. Halphen proved (cf. *Traité des Fonctions Elliptiques*, vol. 2, 1888, p. 128) that the difference in longitude between successive points of tangency to limiting circles,is always less than π; and V. Puisseux proved (cf. Journal de Mathématiques, vol. 7 (1842), p. 517) that this difference in longitude is always greater than $\pi/2$. Simple proofs of both theorems are given by A. de Saint Germain in the Bulletin des Sciences Mathématiques, 1896, 1898, 1901, and in the Mémoires de l'Académie de Caen, 1901. Cf. also the remark of Kellogg (loc. cit., p. 521) concerning this change in longitude in the case of the gyroscope.

Furthermore, by choosing a sufficiently small (or γ sufficiently large) we may make $\Delta\psi(0)$ differ from $\pi/2$ by an amount which can be made as small as desired. Thus by first choosing a small and then choosing $h<0$ and $|h|$ small, we can make $\Delta\psi(h)$ differ from zero by an amount which can be made arbitrarily small. When $h<0$, we recall that Γ is a curve with loops.

At the south pole, $\psi_0=-\gamma/(1-u_0)$, and the change in longitude, $\Delta\psi$, is given by the equation

$$\Delta\psi = \int_{-1}^{u_0} \frac{-\gamma du}{(1-u)[a(u_0-u)(u+1)(u_2-u)]^{1/2}}.$$

The evaluation of this integral gives the expression

$$\Delta\psi = \frac{-\gamma\pi}{[2a(1-u_0)(u_2-\bar{u})]^{1/2}} \quad \text{where } -1<\bar{u}<u_0.$$

But, since $u_2=1+2\gamma^2/[a(1-u_0)]$, we obtain the equation

$$\Delta\psi = -\frac{\pi}{2}\frac{1}{\left(1+\dfrac{a(1-u_0)(1-\bar{u})}{2\gamma}\right)^{1/2}}.$$

Hence $-\pi/2<\Delta\psi<0$, and by a proper choice of a (or γ) $\Delta\psi$ may be made to differ from 0 or $-\pi/2$ by an arbitrarily small amount.

If we let $\psi_0=-\gamma/(1-u_0)+h$, then corresponding to any value of h, we obtain a path $\Gamma(h)$. If $|h|$ is small and $h>0$, $\Gamma(h)$ is a curve with loops, and if $h<0$, $\Gamma(h)$ is a wavy curve westward. This situation is shown in Fig. XIV.

It is easily seen that at the south pole*

$$\lim_{h\to 0^+}\Delta\psi(h) = \Delta\psi(0)+\pi/2$$

and

$$\lim_{h\to 0}\Delta\psi(h) = \Delta\psi(0)-\pi/2.$$

But, since $-\pi/2<\Delta\psi(0)$, it follows that when $|h|$ is sufficiently small

$$0<\Delta\psi(h)<\pi/2 \quad \text{when } h>0$$

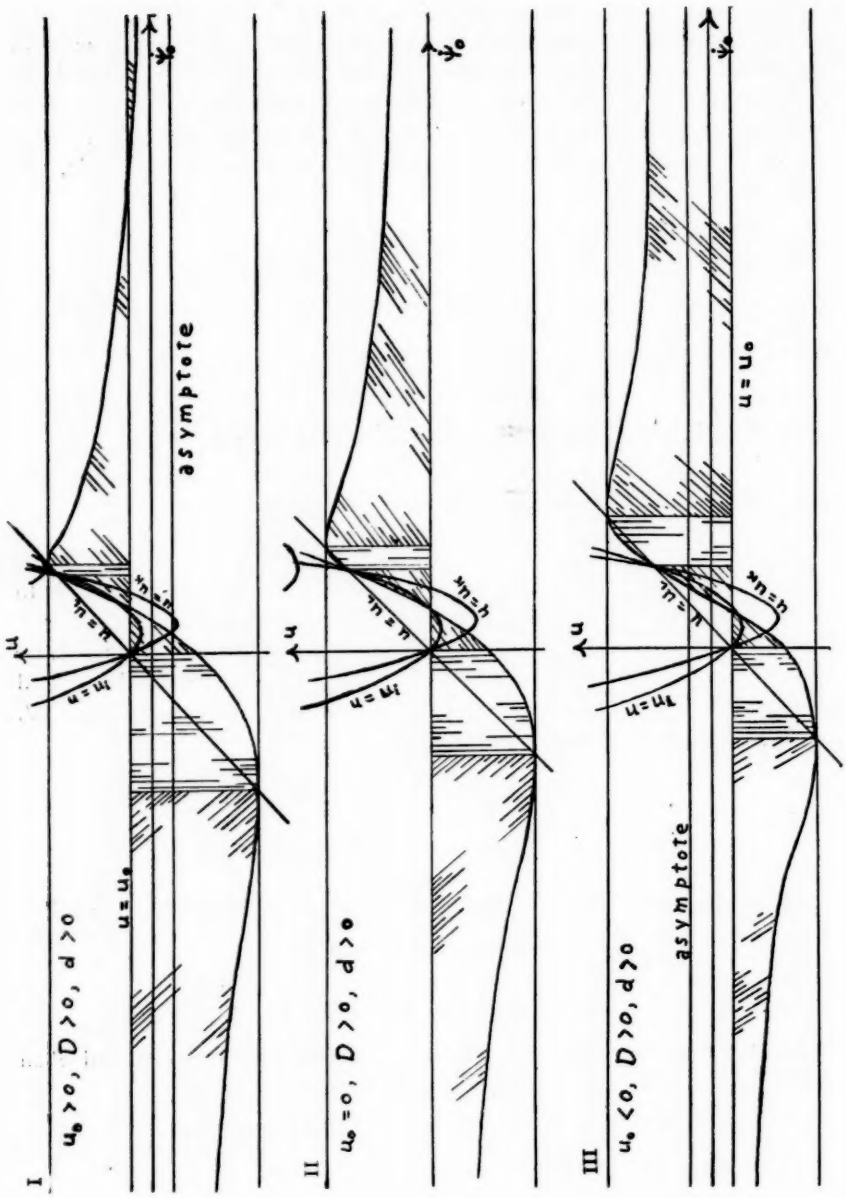
and

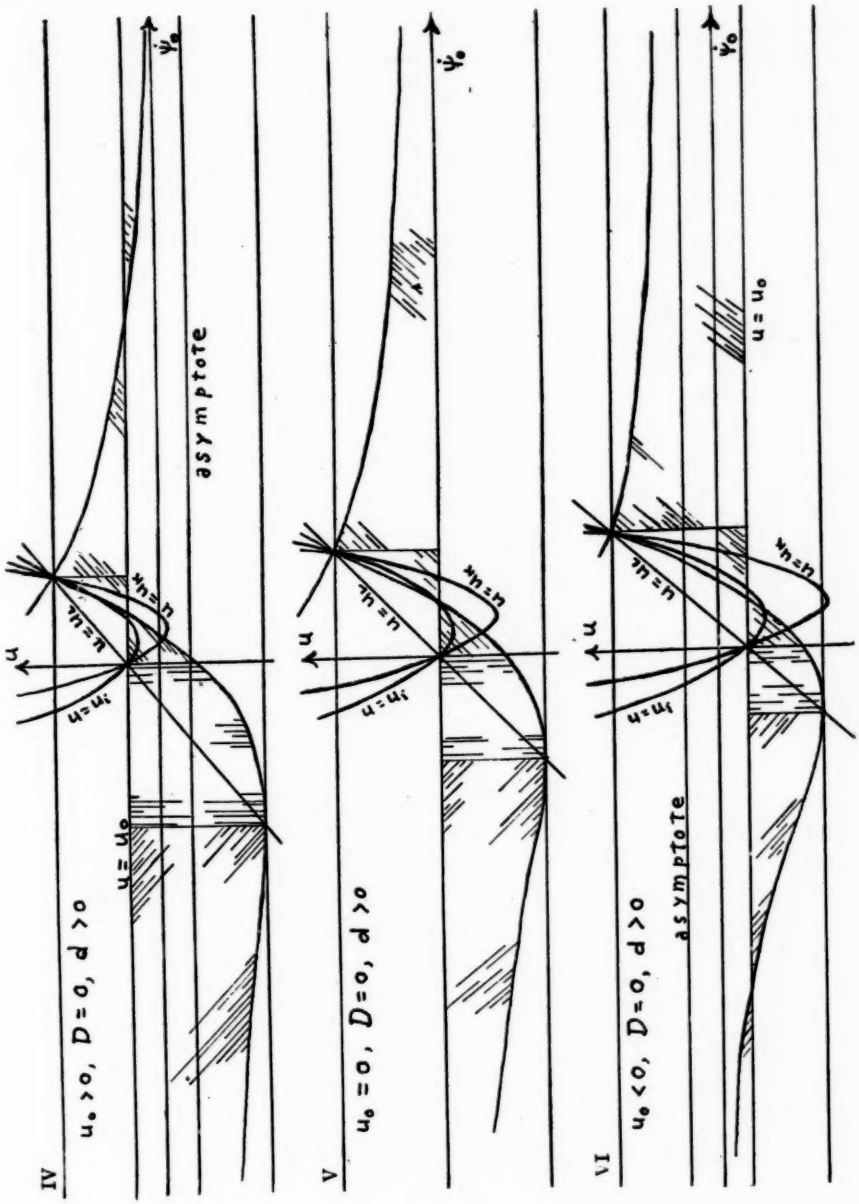
$$-\pi<\Delta\psi(h)<-\pi/2 \quad \text{when } h<0.$$

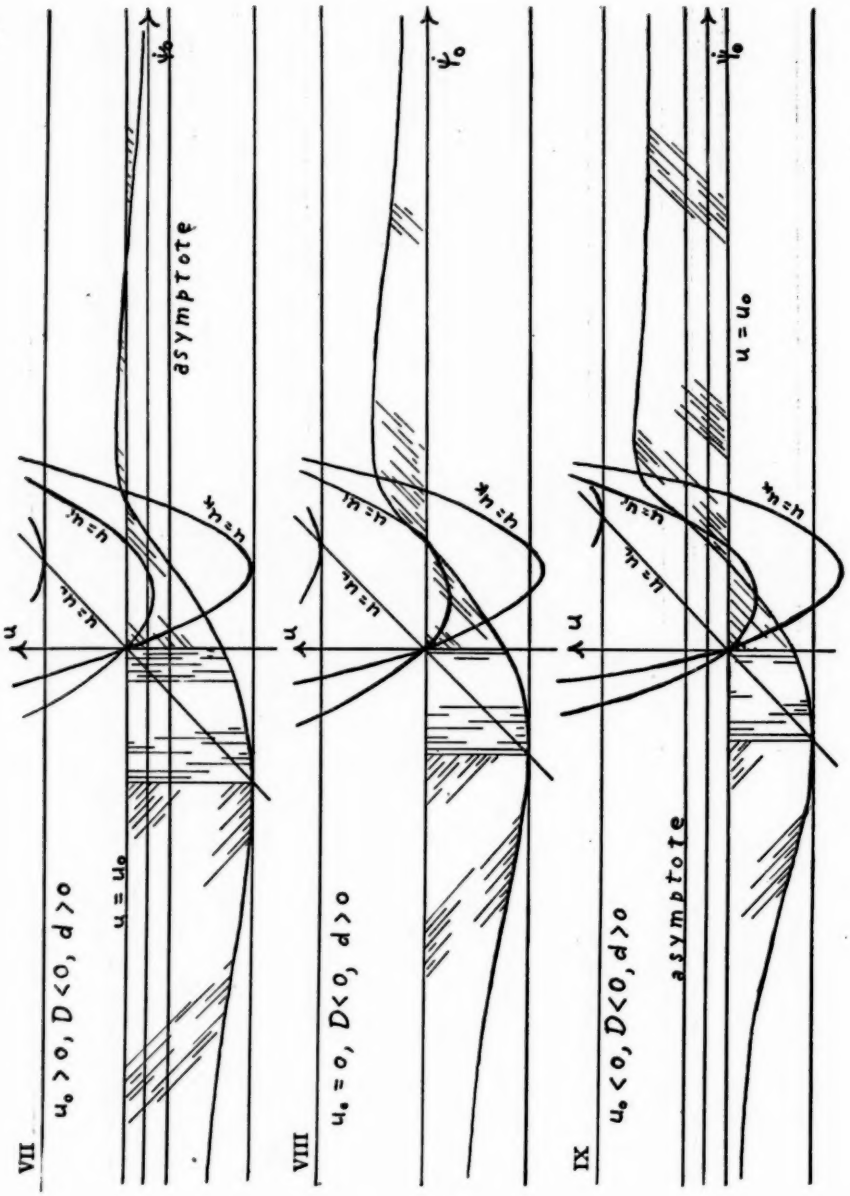
When $h>0$, $\Delta\psi(h)$ can be made arbitrarily near to either 0 or $\pi/2$, and when $h<0$, $\Delta\psi(h)$ can be made arbitrarily near to either $-\pi$ or $-\pi/2$.

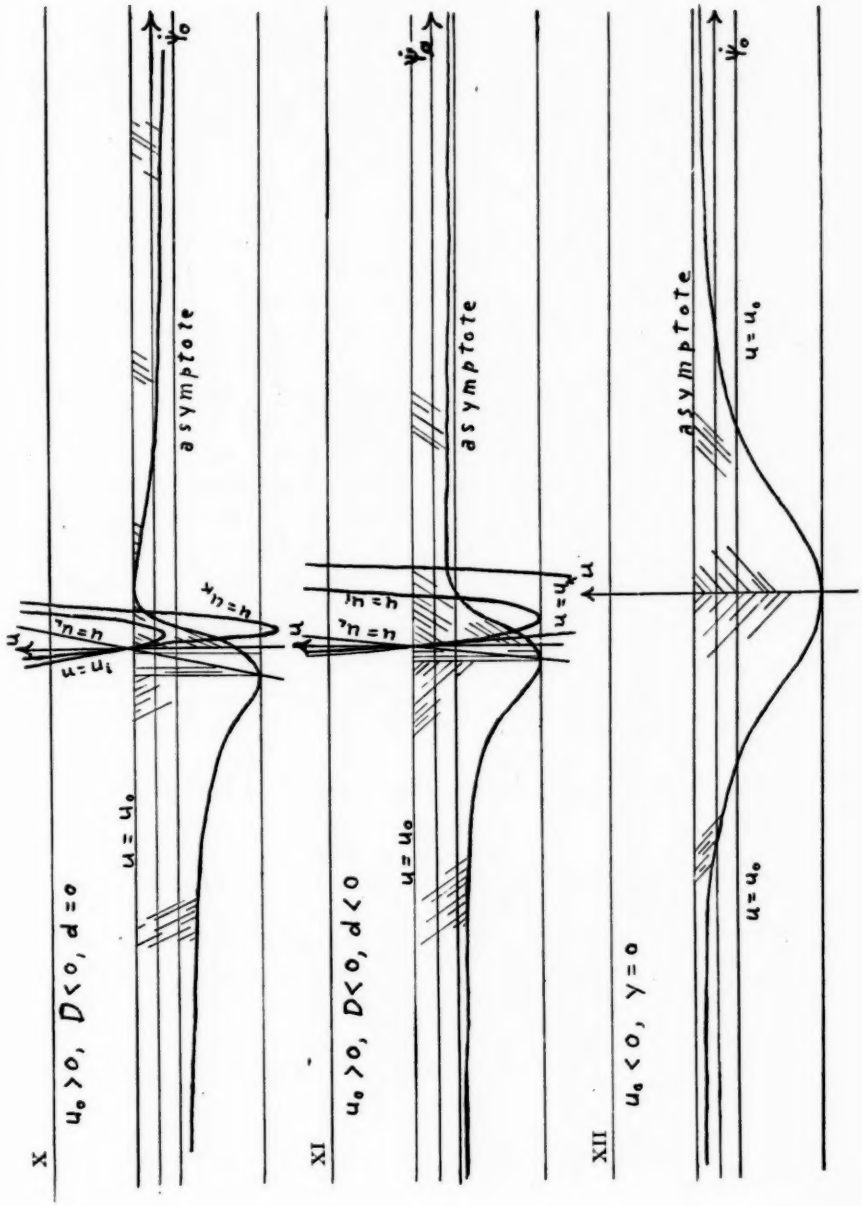
* These equations may also be obtained analytically. Cf. footnote p. 757.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

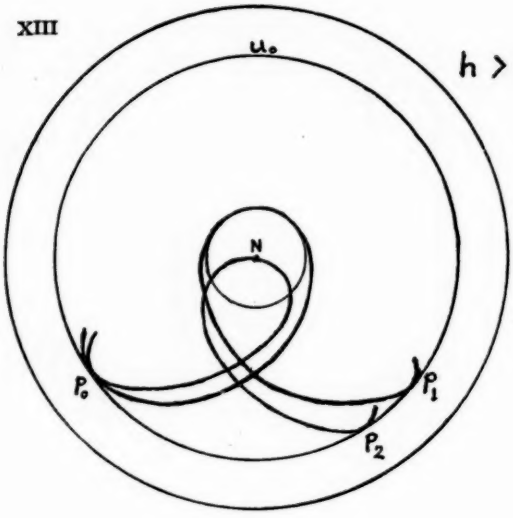

I
$u_0 > 0,\ D > 0,\ d > 0$
asymptote
$u = u_0$
$\eta = \eta_1$
$\eta = \eta_2$
$u = u_1$
$u = u_2$
II
$u_0 = 0,\ D > 0,\ d > 0$
$u = u_1$
$u = u_2$
$\eta = \eta_1$
$\eta = \eta_2$
III
$u_0 < 0,\ D > 0,\ d > 0$
asymptote
$u = u_0$
$\eta = \eta_1$
$\eta = \eta_2$
$u = u_1$
$u = u_2$


IV
$u_0 > 0$, $D = 0$, $d > 0$
asymptote
$u = u_0$
$u = u_k$
$u = u_h$
$\dot{\psi}_0$
V
$u_0 = 0$, $D = 0$, $d > 0$
$u = u_k$
$u = u_h$
$\dot{\psi}_0$
VI
$u_0 < 0$, $D = 0$, $d > 0$
asymptote
$u = u_0$
$u = u_k$
$u = u_h$
$\dot{\psi}_0$

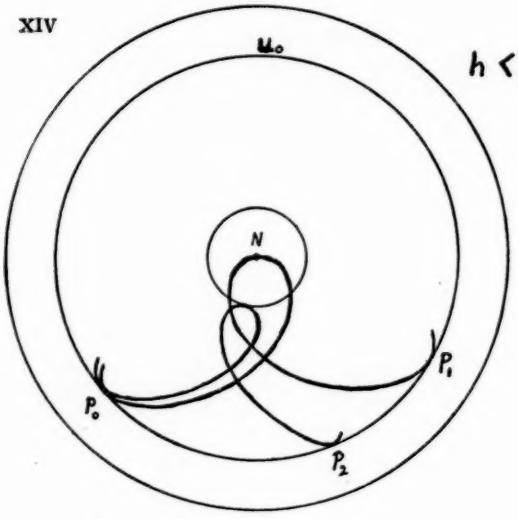

VII
$u_0 > 0, D < 0, d > 0$
$u = u_0$
asymptote
$u = u_1$
$u = u_2$
$u = u_3$
u
ψ_0
VIII
$u_0 = 0, D < 0, d > 0$
$u = u_1$
$u = u_2$
$u = u_3$
u
ψ_0
IX
$u_0 < 0, D < 0, d > 0$
asymptote
$u = u_0$
$u = u_1$
$u = u_2$
$u = u_3$
u
ψ_0


X
$u_0 > 0$, $D < 0$, $d = 0$
asymptote
$u = u_0$
$u = u_1$
$u = u_2$
$u = u_3$
$\dot{\psi}_0$
XI
$u_0 > 0$, $D < 0$, $d < 0$
asymptote
$u = u_0$
$u = u_1$
$u = u_2$
$u = u_3$
$\dot{\psi}_0$
XII
$u_0 < 0$, $\gamma = 0$
asymptote
$u = u_0$
$u = u_0$
$\dot{\psi}_0$
u

XIII



XIV



THE INVARIANT INTEGRAL AND THE INVERSE PROBLEM IN THE CALCULUS OF VARIATIONS*

BY

THOMAS H. RAWLES†

INTRODUCTION

1. The simplest problem in the calculus of variations is the determination of that curve, $y=Y(x)$, joining two points, P_1 and P_2, in the xy plane along which the definite integral

$$J = \int_{x_1}^{x_2} f(x, y(x), y'(x))dx$$

must be taken in order that the value of the integral may be an extremum. The differential equation which defines the desired curve is of the second order and its general solution is a two-parameter family of curves known as the extremals of the problem. By the successive substitution of the coördinates of the end points, P_1 and P_2, in the equation of the extremals we obtain two equations for the determination of the values of the parameters in terms of these coördinates.

2. The inverse problem is, as its name suggests, that of determining the form of the integrand function which has as its extremals a given two-parameter family of curves.

3. Besides the simplest problem as we have stated it there is the problem of determining the extremals of an integral,

$$J = \int f(x, y_1(x), y_2(x), \cdots, y_n(x), y_1'(x), y_2'(x), \cdots, y_n'(x))dx,$$

whose integrand function contains n dependent variables and their derivatives. The extremal in this case is a curve in a hyperspace of $n+1$ dimensions and is defined by the intersection of n hypersurfaces. The problem is of order $2n$ and the constants are determined by successively substituting the coördinates of the two end points in the n equations which define the extremals.

---

* Presented to the Society, December 28, 1927; received by the editors June 24, 1927.

† Sterling Fellow in Mathematics, Yale University, 1926–1927.

Corresponding to this there is of course an inverse problem in which the extremals are given by a family of curves involving $2n$ parameters in a hyperspace of $n+1$ dimensions.

4. The inverse problem in the calculus of variations was first discussed by Darboux* who was led to its study in connection with his researches in the theory of surfaces. He was able to apply his results to certain problems which had been previously investigated by Beltrami and Dini. He also solved the problem of finding the form of the integrand function which has families of straight lines as its extremals.

5. The first application of Darboux's ideas was that of Hamel,† who investigated the geometries in which it is assumed among other axioms that the shortest distance between two points is measured along a straight line joining them. He first developed the most general form of a line element of such a geometry in a plane. This portion of Hamel's paper contains nothing but what is a direct consequence of Darboux's work. He found in addition the form of the integrand function which has as its extremals straight lines in space.

6. Employing methods similar to those of Hamel, Stromquist‡ was able to obtain the most general form of line element of a geometry in which the shortest distance between two points is measured along the segment of a circle which joins them and has its center on the boundary of the half plane. He then introduced the restriction that the transversals are perpendicular to the extremals and showed that this leads to a unique form of line element.

7. Miles§ obtained the integrand functions which have as extremals certain two-parameter families of conics. This investigation as well as those of Hamel and Stromquist is a direct application of Darboux's methods.

8. By using as the independent variable the angle between the tangent and the positive direction of the x axis Bliss‖ was able to discuss the problem without restricting himself to curves whose tangents do not become parallel to the axis of the dependent variable. In order to determine the integrand function uniquely he assigned a particular family of transversals to any one-parameter family of extremals which form a field.

9. Since the methods developed by Darboux and Bliss give an infinite number of solutions a further condition must be imposed in order to give

* Darboux, *Théorie des Surfaces*, vol. III, §§ 604, 605.

† Hamel, Mathematische Annalen, vol. 57 (1903), p. 231; also his dissertation, Göttingen, 1901.

‡ Stromquist, these Transactions, vol. 7 (1906), p. 181.

§ Miles, American Mathematical Monthly, vol. 20 (1913), p. 117.

‖ Bliss, Annals of Mathematics, (2), vol. 9 (1907), p. 127.

uniqueness to the result. The condition which naturally suggests itself is one which relates the slope of the extremals to that of an associated family of curves known as the transversals.* When such a condition is imposed the determination of the integrand function becomes unique except for a constant factor which has no significance as far as the first variation is concerned.

10. The objects of this paper are as follows:

I. To determine the general form which the integrand function must have in order that there may exist a prescribed relation between the slope of the extremal and that of the normal to the transversal.

II. Having found the general form of the integrand function, to determine the particular function which has as its extremals a given family of curves.

The cases of one and n dependent variables will be given separate discussion with the same ends in view in order to bring out certain conditions which must be satisfied when more than one dependent variable is present.

The problem will be attacked by means of the Hilbert invariant integral.

I. The case of one dependent variable

1. **The field of extremals.** A one-parameter family of extremals, $y=Y(x, a)$, is said to form a field† when in a region which we may define by $a_1 \leqq a \leqq a_2$, and $x_1 \leqq x \leqq x_2$, no two curves of the family have a point in common. This implies of course that for any pair of values of x and y in such a region the equation $y=Y(x, a)$ has one and but one root in a. Therefore we may define a function $a=A(x, y)$, known as the inverse function of the field. Since the extremals of the field have continuous first derivatives there is a definite slope associated with each point of the field which we shall denote by

$$p(x,y) = Y_x(x,A(x,y)), \tag{1}$$

calling it the slope function of the field.

An important theorem on the existence of a field is the following.‡ Suppose that a family of extremals, $y=Y(x, a)$, contains the curve C_0 for $a=a_0$, and that the functions Y and Y_x are of class C' in the region which satisfies the inequalities $X_1 \leqq x \leqq X_2$, and $|a-a_0|<d$; and also that $X_1<x_1<x_2<X_2$. Then if $Y_a(x, a)\neq 0$ in the open interval (x_1, x_2), a quantity k may be chosen so small that the family of curves $y=Y(x, a)$ forms a field about the curve C_0 when $x_1 \leqq x \leqq x_2$, and $|a-a_0|<k$.

* Bliss, loc. cit., p. 134; Stromquist, loc. cit., p. 180.

† Bolza, *Vorlesungen über Variationsrechnung*, Berlin, 1909, p. 96.

‡ Bolza, *Vorlesungen*, p. 100.

2. **The invariant integral.** It was shown by Hilbert† that the integral

$$J^* = \int_{P_0}^{P} f(x,y,p(x,y)) + (y' - p(x,y))f_{y'}(x,y,p(x,y))dx, \tag{2}$$

taken along any curve lying wholly within the field, is independent of the path of integration and depends only upon the coördinates of the end points. If, therefore, $J^*(P_0, P)$ represents the integral taken from a point P_0 to any other point in the field it will define a function,

$$J^*(P_0,P) = w(x,y). \tag{3}$$

It is known further‡ that the derivatives of this function are

$$\begin{aligned} \frac{\partial w}{\partial x} &= f(x,y,p(x,y)) - p(x,y)f_{y'}(x,y,p(x,y)), \\ \frac{\partial w}{\partial y} &= f_{y'}(x,y,p(x,y)). \end{aligned} \tag{4}$$

The family of curves $w=$constant are known as the transversals of the field. They are of class C'', and along any one of them the invariant integral has a constant value. We shall call w the transversal function of the field.

If from equations (4) we eliminate p, a partial differential equation in w is obtained. Then, if there can be found in any way a solution of this equation containing a non-linear constant, α, the extremals are given by

$$\frac{\partial w(x,y,\alpha)}{\partial \alpha} = c, \tag{5}$$

where c is the second arbitrary constant which the solution of the Euler equation must contain.

3. Having in mind these properties of the field, and of the transversal function associated with it, we shall now seek the form which the integrand function must assume in order that a given relation between the slope of the extremal and that of the normal to the transversal shall be satisfied.

Let us suppose that the extremals are given by equations (5), and that the slope of the transversals of the field is related to that of the extremals by a given function of the slope of the extremals and of the coördinates of the point,

$$w_y = m(x,y,p)w_x. \tag{6}$$

† Bolza, *Vorlesungen*, p. 106.

‡ Bolza, *Vorlesungen*, p. 129.

We shall define as a transversality any relation such as (6) in which the function m is of the class C', excepting only the function defined by $1+mp=0$.

Now, if we differentiate (5) with respect to x the result is

$$w_{\alpha x} + w_{\alpha y}y' = w_{\alpha x} + w_{\alpha y}p = 0. \tag{7}$$

This serves to define p, the slope function of the field.

In order to proceed with the discussion of the problem we shall fix our attention on that portion of the field for which $1+mp\neq 0$ and $w_x\neq 0$. Differentiating (6) with respect to α we obtain

$$\frac{w_{y\alpha}}{w_x} - \frac{w_y w_{x\alpha}}{w_x^2} = m_\alpha.$$

This becomes

$$-\frac{1}{p}\frac{w_{x\alpha}}{w_x} - m\frac{w_{x\alpha}}{w_x} = m_\alpha$$

when we replace w_y and $w_{y\alpha}$ by their values which may be obtained from (6) and (7) respectively.

Solving this last equation for $w_{x\alpha}/w_x$ we find

$$\frac{w_{x\alpha}}{w_x} = -\frac{m_\alpha p}{1+mp},$$

or, after adding and subtracting the term $mp_\alpha/(1+mp)$,

$$\frac{w_{x\alpha}}{w_x} = -\frac{\partial}{\partial\alpha}\log(1+mp) + \frac{mp_\alpha}{1+mp}.$$

Upon integrating this we now obtain

$$w_x = g(x,y)\left(\frac{1}{1+mp}\right)e^{\Theta}, \tag{8}$$

where

$$\Theta = \int \frac{m\,dp}{1+mp}, \tag{9}$$

and $g(x, y)$ is an arbitrary function of its arguments.

By means of equations (6) and (8) we at once arrive at

$$w_y = g(x,y)\left(\frac{m}{1+mp}\right)e^{\Theta}. \tag{10}$$

We now substitute these expressions for w_x and w_y in (4) and find

$$f(x,y,p) = w_x + pw_y = g(x,y)e^{\Theta}. \tag{11}$$

This result was first given by Stromquist* in 1907. The method by which we have derived it, employing as it does the invariant integral, is quite different from that which he used.

4. Having found the general form of the integrand function for a specified transversality, we now suppose that, in addition, a two-parameter family of curves, $\phi(x, y, a, c)=0$, is given and we seek to determine the particular integrand function for which they are extremals.

Assuming that in a region in the xy plane, and for certain values of a and c, ϕ_c is not equal to zero, it follows that when a is fixed and c is given nearby values the family of extremals thus obtained forms a field. This is a consequence of the theorem stated in the first paragraph of this section.

Moreover, we can solve for c, finding

$$C(x,y,a) = c. \tag{12}$$

Then we may differentiate with respect to x and obtain

$$p = -\frac{C_x}{C_y}, \tag{13}$$

where p is the slope function of the field.

The existence of a field implies the existence of a family of transversals, $w(x, y, \alpha)=$ constant, from which we can derive the extremals by differentiation with respect to α. The extremals would be given by

$$w_\alpha(x,y,\alpha) = \gamma. \tag{14}$$

Now, when we differentiate this with respect to x, we find

$$p = -\frac{w_{\alpha x}}{w_{\alpha y}}, \tag{15}$$

where p is again the slope function of the field. Since the expression obtained from (15) is identical with that obtained from (13) it follows that a must be some function of α.

Since w is of class C'' we must have

$$w_{xy} = w_{yx}.$$

Therefore, it follows that

$$\frac{\partial}{\partial y}\left(\frac{ge^{\Theta}}{1+mp}\right) = \frac{\partial}{\partial x}\left(\frac{mge^{\Theta}}{1+mp}\right).$$

* Stromquist, Annals of Mathematics, (2), vol. 9 (1907), p. 57.

This may be written

$$\frac{mg_x}{g} - \frac{g_y}{g} = \frac{\partial}{\partial y}(\Theta - \log(1 + mp)) - m_x - m\frac{\partial}{\partial x}(\Theta - \log(1 + mp)) = \Phi(x, y, \alpha). \tag{16}$$

This is a partial differential equation for the determination of $g(x, y)$. To obtain a solution we make use of the fact that g is independent of α while it can be shown that m must contain α. For, if m were independent of α we would have

$$w_{y\alpha} - mw_{x\alpha} = 0,$$

from (6). But by (15)

$$w_{x\alpha} + pw_{y\alpha} = 0.$$

If the determinant of the system formed by these two equations is to vanish we must have

$$1 + mp = 0,$$

which is contrary to an assumption which we have made.

Differentiating (16) with respect to α, or, what amounts to the same thing, with respect to a, and dividing by $-m_a$ we obtain

$$\frac{g_x}{g} = \frac{1}{m_a}\Phi_a. \tag{17}$$

Substituting this result in (16) we find at once

$$\frac{g_y}{g} = \frac{m}{m_a}\Phi_a - \Phi. \tag{18}$$

Knowing the partial derivatives of $\log g$ we can easily find g by integration.

It is not difficult to show that the solutions thus obtained are unique except for a constant factor, which, as we have already remarked, has no significance as far as the first variation is concerned. For, if there were two solutions g and g', we would have

$$\left(\frac{g_y}{g} - \frac{g'_y}{g}\right) - m\left(\frac{g_y}{g} - \frac{g'_x}{g}\right) = 0,$$

which may be written

$$\frac{\partial}{\partial y}\log\frac{g}{g'} - m\frac{\partial}{\partial x}\log\frac{g}{g'} = 0.$$

Since g and g' are independent of α, and m is not, this is impossible unless g/g' is a constant.

Example I. Let us assume that the prescribed transversality is that the transversals shall intersect the extremals orthogonally. If this is the case then $m=p$ and

$$\Theta = \int \frac{m\,dp}{1+mp} = \int \frac{p\,dp}{1+p^2} = \frac{1}{2}\log(1+p^2).$$

The integrand function, therefore, is of the form

$$f = g(x,y)(1+p^2)^{1/2}.$$

Further, let us suppose that the prescribed family of curves which are to be the extremals of the problem is the family of parabolas given by

$$y = a + \frac{(x+c)^2}{4a}.$$

First, we notice that we can solve this equation for c. After differentiating with respect to x we calculate p, and then Φ by substitution in (16). The result is

$$p = \left(\frac{y-a}{a}\right)^{1/2},$$

$$\Phi = -\frac{1}{2}\frac{\partial}{\partial y}\log(1+p^2) - p_x + \frac{p}{2}\frac{\partial}{\partial x}\log(1+p^2) \tag{16'}$$

$$= -\frac{1}{2}\frac{\partial}{\partial y}[\log y - \log a] - \frac{\partial}{\partial x}\left(\frac{y-a}{a}\right)^{1/2}$$

$$+\frac{1}{2}\left(\frac{y-a}{a}\right)^{1/2}\frac{\partial}{\partial x}\log\frac{y}{a} = -\frac{1}{2y}.$$

On substituting these results in (17) and (18) we see that

$$\frac{g_x}{g} = 0, \quad \frac{g_y}{g} = \frac{1}{2y}.$$

Hence $g=$ constant $y^{1/2}$.

On the other hand we might wish to solve for a. In this case we would obtain

$$a = \frac{y \pm (y^2-(x+c)^2)^{1/2}}{2}.$$

Differentiating this with respect to x we then find

$$p = \frac{x + c_1}{x \pm [y - (x + c)^2]^{1/2}}.$$

When this result is substituted in (16′) that equation becomes

$$\begin{aligned}\Phi &= -\frac{1}{2}\left(\frac{1}{y} - (\pm)\frac{1}{(y^2 - (x + c)^2)^{1/2}}\right) \\ &\quad - \frac{(y \pm (y^2 - (x + c)^2)^{1/2})(y^2 - (x + c)^2)^{1/2} \pm \frac{1}{2}(x + c)^2}{(y \pm (y^2 - (x + c)^2)^{1/2})^2(y^2 - (x + c)^2)^{1/2}} \\ &= -\frac{1}{2y} - \frac{-\frac{1}{2}(\pm)(y \pm (y^2 - (x+c)^2)^{1/2} \pm (y^2 - \frac{1}{2}(x+c)^2) + y(y^2 - (x+c)^2)^{1/2}}{(y^2 - (x + c)^2)^{1/2}(y \pm (y^2 - (x + c)^2)^{1/2})^2} \\ &= -\frac{1}{2y}\end{aligned}$$

as before.

5. While every two-parameter family of curves may be regarded as the extremals of a problem in the calculus of variations it does not follow that any family may serve as the extremals of a problem with a prescribed transversality. The necessary conditions to be satisfied in order that it shall do so are easily obtained.

Since the left member of (17) is independent of a the right member must also be independent. From this it follows that

$$\Phi_{aa} - \frac{m_{aa}}{m_a}\Phi_a = 0. \tag{19}$$

When the right member of (17) is independent of a the right member of (18) is likewise. For differentiating the right member of (18) we have

$$\frac{m}{m_a}\left(\Phi_{aa} - \frac{m_{aa}}{m_a}\Phi_a\right) - \Phi_a + \frac{m_a}{m_a}\Phi_a = \frac{m}{m_a}\left(\Phi_{aa} - \frac{m_{aa}}{m_a}\Phi_a\right).$$

This vanishes on account of (19).

A second condition to be satisfied is that equations (17) and (18) shall actually define a function. This requires that

$$\frac{\partial}{\partial x}\left(\Phi + \frac{m}{m_a}\Phi_a\right) = \frac{\partial}{\partial y}\left(\frac{1}{m_a}\Phi_a\right). \tag{20}$$

These conditions are also sufficient. For when they are satisfied there exists a function, w, whose derivatives are given by

$$w_x = \frac{ge^{\Theta}}{1 + mp},$$

and

$$w_y = \frac{mge^{\Theta}}{1 + mp}.$$

Since these derivatives satisfy (16), which may be written

$$\frac{\partial}{\partial y}(f - pf_p) = \frac{\partial}{\partial x}f_p,$$

and since any function p which does satisfy this equation is a slope function of a field of extremals, our given curves are actually the extremals of the integrand function which we have found.*

Example II. A two-parameter family of curves which does not satisfy (19) for the orthogonal transversality, $m = p$, is $y = a(x-c)^2$. Here

$$p = 2(ay)^{1/2}, \quad \Phi = \frac{2}{(1/a) + 4y}.$$

Condition (19) becomes

$$\frac{p_{aa}}{p_a}\Phi_a - \Phi_{aa} = -\frac{1}{2a}\frac{2}{(1 + 4ay)^2} + \frac{16y}{(1 + 4ay)^3} \not\equiv 0.$$

Example III. A two-parameter family in which (19) is satisfied but (20) is not, is the following:

$$y - \int_0^x \tan(x^2 - a)dx = c,$$

$$p = \tan(x^2 - a), \quad \Phi = -2x.$$

Condition (20) becomes

$$\frac{\partial}{\partial x}(-2x) = \frac{\partial}{\partial y}(0),$$

which is not satisfied.

II. The case of n dependent variables

1. Analogous to the transversal lines in the case of one dependent variable and to the transversal surfaces in the case of two, we have here to deal with a family of transversal hypersurfaces in a space of $n+1$ dimensions. These we shall indicate by $w(x, y_1, y_2, \cdots, y_n, \alpha_1, \alpha_2, \cdots, \alpha_n) = c$, or, more briefly, by $w(x, y, \alpha) = c$. In this expression the y indicates the presence of the n coördinates and α that of the n constants.

* Bolza, *Vorlesungen*, p. 106.

The derivatives of the transversal function are known to be

$$w_x = f(x,y,p) - \sum p_i f_{p_i}, \quad w_{y_i} = f_{p_i} \qquad (i = 1,2,\cdots,n). \tag{1}$$

These functions are formed by replacing the derivatives in the integrand function by the slope functions of the field. Along an extremal we have, of course,

$$\frac{dy_i}{dx} = p_i(x,y,\alpha).$$

The n equations which define the extremals may be derived from the transversal function by differentiation with respect to the constants.* They are

$$w_\alpha = (x,y,\alpha) = \gamma \qquad (i = 1,2,\cdots,n). \tag{2}$$

The slope functions of the field can be obtained by differentiating these equations with respect to x.

We obtain the n equations

$$w_{\alpha_i x} + \sum p_j w_{\alpha_i y_j} = 0 \qquad (i = 1,2,\cdots,n). \tag{3}$$

These can be solved for the p_i since we are not concerned with regions for which the determinant of this system vanishes. In fact, one of the restrictions introduced in the direct problem is that the derivatives must remain finite.

2. A transversality for n dependent variables will be defined as a set of n relations

$$w_{y_i} = m_i w_x, \tag{4}$$

which satisfy the inequality

$$1 + \sum p_i m_i \neq 0, \tag{5a}$$

and the $n(n-1)/2$ equations

$$\sum p_k\left(m_k\frac{\partial m_i}{\partial p_j} - m_i\frac{\partial m_k}{\partial p_j}\right) + \frac{\partial m_i}{\partial p_j} = \sum p_k\left(m_k\frac{\partial m_j}{\partial p_i} - m_j\frac{\partial m_k}{\partial p_i}\right) + \frac{\partial m_j}{\partial p_i}. \tag{5b}$$

By substituting in these conditions the derivatives of the transversal function we shall show that these conditions are always satisfied in any problem.

From (1) and (4) it follows that

$$m_i = \frac{f_{p_i}}{f - \sum f_{p_j} p_j}.$$

* Bolza, *Vorlesungen*, p. 601.

If the inequality (5a) were not satisfied we would have

$$1+\frac{\sum f_{p_i}p_j}{f-\sum f_{p_j}p_j}\equiv 0,$$

and, therefore, $f(x, y, p)\equiv 0$.

To show that the $n(n-1)/2$ equations (5b) are always satisfied we substitute in them the derivatives of the transversal function. In order to introduce logarithms we first divide by m_im_j. The left member of (5b) then becomes

$$\begin{aligned}
&\frac{1}{m_j}\left[\sum p_km_k\frac{\partial}{\partial p_j}\log\frac{m_i}{m_k}+\frac{\partial}{\partial p_j}\log m_i\right]\\
&=\frac{f-\sum f_{p_k}p_k}{f_{p_j}}\left[\frac{\sum p_kf_{p_k}}{f-\sum f_{p_k}p_k}\left(\frac{f_{p_ip_j}}{f_{p_i}}-\frac{f_{p_kp_j}}{f_{p_k}}\right)+\frac{f_{p_ip_j}}{f_{p_i}}\right.\\
&\qquad\left.-\frac{f_{p_j}-f_{p_j}-\sum f_{p_kp_j}p_k}{f-\sum f_{p_k}p_k}\right]\\
&=\frac{1}{f_{p_j}}\left[\frac{f_{p_ip_j}}{f_{p_i}}\sum f_{p_k}p_k-\sum f_{p_kp_j}p_k\right.\\
&\qquad\left.+\frac{f_{p_ip_j}}{f_{p_i}}(f-\sum f_{p_k}p_k)+\sum f_{p_kp_j}p_k\right].
\end{aligned}$$

In this last expression the summations cancel and we have, finally,

$$ff_{p_ip_j}/f_{p_i}f_{p_j}.$$

Since this is symmetrical in the derivatives with respect to p_i and p_j it follows that any transversality which arises in a problem necessarily satisfies equations (5b).*

* It is significant that if relations (5a) and (5b) hold for any set of variables, y_i, they also hold for any other set, $z_i=Z_i(y)$. Employing the tensor notation for the moment we see that

$$p^i=\frac{\partial y_i}{\partial z_k}q^k,\qquad m_i=\frac{\partial z_j}{\partial y_i}n_j,$$

where $q^k=dz_k/dx$ and $n_j=(\partial w/\partial z_j)/(\partial w/\partial x)$ are the components of the tensor in the new variables. By means of these expressions we see that (5a) becomes

$$1+p^im_i=1+\frac{\partial y_i}{\partial z_k}\frac{\partial z_j}{\partial y_i}q^kn_j=1+q^in_j.$$

Hence this quantity is an invariant.

The equations (5b) are tensor equations formed from the components of a covariant tensor of rank 2. In addition to the relations already mentioned we have

$$R_{ij}=S_{rs}\frac{\partial z_r}{\partial y_i}\frac{\partial z_s}{\partial y_j},$$

3. We now differentiate equations (4) with respect to one of the constants and obtain

$$\frac{1}{w_x}\frac{\partial w_{y_i}}{\partial \alpha_j} - \frac{w_{y_i}}{w_x^2}\frac{\partial w_x}{\partial \alpha_j} = \frac{\partial m}{\partial \alpha_j} \qquad (i = 1,2, \cdots, n).$$

On replacing w_{y_i} by $m_i w_x$, according to (4), these become

$$\frac{1}{w_x}\frac{\partial w_{y_i}}{\partial \alpha_j} - \frac{m_i}{w_x}\frac{\partial w_x}{\partial \alpha_j} = \frac{\partial m_i}{\partial \alpha_j}.$$

Now we multiply each of these n equations by $-p_i$ and add, with the result

$$-\frac{1}{w_x}\sum p_i \frac{\partial w_{y_i}}{\partial \alpha_j} + \frac{1}{w_x}\frac{\partial w_x}{\partial \alpha_j}\sum m_i p_i = -\sum p_i \frac{\partial m_i}{\partial \alpha_j}.$$

Substituting $\partial w_x/\partial \alpha_j$ for the first summation in the left member, by (3), the last equation reduces to

$$\frac{1}{w_x}(1 + \sum m_i p_i)\frac{\partial w_x}{\partial \alpha_j} = -\sum p_i \frac{\partial m_i}{\partial \alpha_j}.$$

Now by adding and subtracting the term $\sum m_i (\partial p_i/\partial \alpha_j)$ in the right member, and then dividing by $1+\sum m_i p_i$, we finally obtain

$$(6) \qquad \frac{1}{w_x}\frac{\partial w_x}{\partial \alpha_j} = -\frac{\partial}{\partial \alpha_j}\log(1 + \sum m_i p_i) + \frac{1}{1 + \sum m_i p_i}\left(\sum m_i \frac{\partial p_i}{\partial \alpha_j}\right).$$

If, as in the case of two variables, δ represents the change in a function when the α_i are varied we may write

$$\delta \log w_x = \sum \frac{1}{w_x}\frac{\partial w_x}{\partial \alpha_j}\delta \alpha_j.$$

Forming the right member of this equation by means of (6) we find

$$(7) \qquad \delta \log w_x = -\delta \log(1 + \sum m_i p_i) + \frac{\sum m_i \delta p_i}{1 + \sum m_i p_i}.$$

The first term on the right is exact. The second is exact if

$$\frac{\partial H_{p_j}}{\partial p_i} = \frac{\partial H_{p_i}}{\partial p_j}$$

---

where $R_{ij} = \partial m_i/\partial p_j$ and $S_{rs} = \partial n_r/\partial q_s$. Upon substituting these expressions in the left members of (5b) the result is, after some reduction,

$$p^k(m_k R_{ij} - m_i R_{kj}) + R_{ij} = [q^r(n_r S_{ab} - n_a S_{rb}) + S_{ab}]\frac{\partial z_a}{\partial y_i}\frac{\partial z_b}{\partial y_j}$$

which displays the tensorial nature of the equations. If T_{ij} represents the left member of (5b) the equations themselves may be written as $T_{ij} - T_{ji} = 0$.

where Hp_i is the coefficient of δp_i. Now this condition reduces to

$$\sum p_k\left(m_k\frac{\partial m_i}{\partial p_j}-m_i\frac{\partial m_k}{\partial p_j}\right)+\frac{\partial m_i}{\partial p_j}-m_im_j=\sum p_k\left(m_k\frac{\partial m_j}{\partial p_i}-m_j\frac{\partial m_k}{\partial p_i}\right)+\frac{\partial m_j}{\partial p_j}-m_jm_i.$$

But this is the same as (5b) when we omit the identical term $-m_im_j$. Since we have supposed (5b) to be satisfied, the right member of (7) is always exact.

On integration we have

$$w_x=\frac{g(x,y)e^{\Theta}}{1+\sum m_ip_i}, \tag{8}$$

where g contains x and the dependent variables y_i but is independent of the α_i, and

$$\Theta=\int\frac{\sum m_i\delta p_i}{1+\sum m_ip_i}. \tag{9}$$

From equations (1) we can eliminate the partial derivatives, f_{p_i}, and find

$$f(x,y,p)=w_x+\sum w_{y_i}p_i.$$

On account of (4) we have

$$w_{y_i}=\frac{m_ig(x,y)e^{\Theta}}{1+\sum m_ip_i}. \tag{8a}$$

Eliminating the derivatives of the transversal function by these last equations we have, finally,

$$f(x,y,p)=g(x,y)e^{\Theta}. \tag{10}$$

4. We have now to show how the function $g(x, y)$ is to be determined when a set of extremals is prescribed in addition to the transversality. Let us suppose, then, that in addition to the transversality a family of extremals involving $2n$ parameters is prescribed, that they have been solved for any n of these parameters, and are given by

$$\psi_1(x,y_1,y_2,\cdots,y_n,a_1,a_2,\cdots,a_n)=\psi_i(x,y,a)=c_i. \tag{11}$$

Since they are to define the extremals we shall suppose that the functions ψ_i are of class C'.

If for a certain set of values of the coördinates the determinant

$$D(\psi)=\frac{\partial(\psi_1,\psi_2,\cdots,\psi_n)}{\partial(y_1,y_2,\cdots,y_n)} \tag{12}$$

does not vanish we can solve for the y_i in terms of x and the constants. We may then substitute these expressions in (11) and write the identities

$$\psi_i(x,y(x,a,c),a)=c_i \qquad (i=1,2,\cdots,n).$$

We can differentiate these identities with respect to c_k, finding

$$\sum \frac{\partial \psi_i}{\partial y_j} \frac{\partial y_j}{\partial c_k} = 0 \quad (i = 1,2, \cdots, n, \text{ except } k),$$

and

$$\sum \frac{\partial \psi_k}{\partial y_j} \frac{\partial \gamma_j}{\partial c_k} = 1.$$

Upon solving these equations for $\partial y_h/\partial c_k$ we obtain

$$\frac{\partial y_h}{\partial c_k} = \frac{C(k,h)}{D(\psi)},$$

where $C(k, h)$ is the cofactor of $\partial \psi_k/\partial y_h$ in $D(\psi)$.

We now form the determinant of the derivatives $\partial y_h/\partial c_k$ and find

$$\frac{\partial(y_1, y_2, \cdots, y_n)}{\partial(c_1, c_2, \cdots, c_n)} = \frac{1}{D(\psi)}$$

by the theorem on reciprocal determinants.* Since $D(\psi)$ is finite on account of the continuity of the functions this determinant is different from zero. Moreover, it is finite in the region for which $D(\psi) \neq 0$. This is a necessary and sufficient condition for the existence of a field of extremals.†

The existence of the field implies the existence of an invariant integral or transversal function containing n constants, α_i, as parameters. If this transversal function were known the extremals could be obtained as usual by differentiation. They would be given by

$$\frac{\partial w(x,y,\alpha)}{\partial \alpha_i} = \gamma_i \qquad (i = 1,2, \cdots, n). \tag{13}$$

The extremals given by equations (13) are the same as those given by (11). It follows that the slope functions of the field, π_i, as defined by

$$\frac{\partial w_x}{\partial \alpha_i} + \sum \pi_j \frac{\partial w_{y_j}}{\partial \alpha_i} = 0 \qquad (i = 1,2, \cdots, n) \tag{14}$$

are identical with those defined by

$$\frac{\partial \psi_i}{\partial x} + \sum p_j \frac{\partial \psi_i}{\partial y_j} = 0 \qquad (i = 1,2, \cdots, n). \tag{15}$$

---

* Weld, *Theory of Determinants*, p. 139.

† Bolza, *Vorlesungen*, p. 635.

In a field the determinants of these systems are different from zero. If they were not, the slope functions would become infinite, which is not the case when a field exists.

Since $p_i(x, y, a) = \pi_i(x, y, \alpha)$ it follows that there must exist transformations such as

$$a_i = A_i(\alpha), \quad \alpha_i = A_i'(a) \qquad (i = 1,2, \cdots, n). \tag{16}$$

This is a consequence of the fact that the n constants a_i are distinct and likewise the n constants α_i.

We shall now show that no one of the functions*

$$m_i(x, y, \pi(x, y, \alpha))$$

can be independent of all the constants. For if it were we would have

$$\frac{\partial m_k}{\partial \alpha_j} = \frac{1}{w_x^2}\left[\frac{\partial^2 w}{\partial \alpha_j \partial y_k}\frac{\partial w}{\partial x} - \frac{\partial w}{\partial y_k}\frac{\partial^2 w}{\partial \alpha_j \partial x}\right] = 0 \qquad (j = 1,2, \cdots, n),$$

and consequently

$$\frac{\partial^2 w}{\partial \alpha_j \partial x} = \frac{\partial^2 w}{\partial \alpha_j \partial y_k}\frac{\partial w}{\partial x}\frac{1}{\partial w/\partial y_k} = \frac{1}{m_k}\frac{\partial^2 w}{\partial \alpha_j \partial y_k}.$$

If we substitute this result in equations (14) they become

$$\left(\pi_k + \frac{1}{m_k}\right)\frac{\partial^2 w}{\partial \alpha_j \partial y_k} + {}_k\sum \pi_i \frac{\partial^2 w}{\partial \alpha_j \partial y_i} = 0 \qquad (j = 1,2, \cdots, n).$$

The subscript k at the left of the summation denotes the omission of that term.

Since we have limited ourselves to a region in which $1+\sum m_i\pi_i \not\equiv 0$ the functions π_i and $\pi_k + 1/m_k$ cannot all be identically zero. Therefore the determinant of the coefficients of the preceding equations must vanish. But this determinant is the same as that of (14) which does not vanish. From this contradiction it follows that m_k is not independent of all the parameters.

If m_k were expressed in terms of $p_i(x, y, a)$ and were independent of all the parameters a_i we would then have

* In this connection it should be pointed out that the problem degenerates if any one of the functions m_i is independent of all the slope functions, p_i. If this were the case we would have $w_{y_i} = m_i w_x$ where m_i is a function of the coördinates only and does not contain the canonical constants of the problem. If a change of variables is made so that the new independent variable, ξ, satisfies the equation, $\xi_{y_i} - m_i\xi_x = 0$, then it is possible to choose the other new variables, η, so as to make one of the derivatives, w_{η_i}, identically zero. This means of course that the derivative $d\eta_i/d\xi$ is absent from the integrand function when expressed in terms of the new variables.

$$\frac{\partial m_k}{\partial a_j} = \sum \frac{\partial m_k}{\partial a_i} \frac{\partial \alpha_i}{\partial a_j} \equiv 0 \qquad (j = 1,2, \cdots, n).$$

Since the derivatives $\partial m_k/\partial \alpha_i$ cannot all be zero the determinant is zero; that is,

$$\frac{\partial(\alpha_1, \alpha_2, \cdots, \alpha_n)}{\partial(a_1, a_2, \cdots, a_n)} \equiv 0.$$

But if this were true we could not obtain the transformations (16). We conclude, therefore, that m_k contains at least one of the constants a_i.

We have found in (8) that

$$w_x = \frac{g(x,y)e^{\Theta}}{1 + \sum m_i p_i}, \tag{8}$$

and in (8a)

$$w_{y_i} = \frac{m_i g(x,y)e^{\Theta}}{1 + \sum m_i p_i}. \tag{8a}$$

Since w is of class C'' the derivatives w_{xy_i} and $w_{y_i x}$ are equal. We express this fact by the n relations

$$\frac{\partial}{\partial y_k}\left(\frac{ge^{\Theta}}{1 + \sum m_i p_i}\right) = \frac{\partial}{\partial x}\left(\frac{m_k g e^{\Theta}}{1 + \sum m_i p_i}\right).$$

These reduce to

$$\begin{aligned} m_k \frac{g_x}{g} - \frac{g_{y_k}}{g} &= \left[\frac{\partial \Theta}{\partial y_k} - \frac{\partial}{\partial y_k} \log (1 + \sum m_i p_i)\right. \\ &= \left.\frac{\partial m_k}{\partial x} - m_k\left(\frac{\partial \Theta}{\partial x} - \frac{\partial}{\partial x} \log (1 + \sum m_i p_i)\right)\right]. \end{aligned} \tag{17}$$

Solutions of the n equations (17) can be obtained as before by using the fact that g and its derivatives are independent of the constants a_i as well as the canonical constants α_i.

If we designate the right members of (17) by Φ_k, then the derivatives of $\log g$ are given by

$$\frac{g_x}{g} = \frac{1}{\partial m_k/\partial a_i} \frac{\partial}{\partial a_i} \Phi_k, \tag{18}$$

$$\frac{g_{y_k}}{g} = \frac{m_k}{\partial m_k/\partial a_i} \frac{\partial}{\partial a_i} \Phi_k - \Phi_k \qquad (i,k = 1,2,3, \cdots, n). \tag{19}$$

From these g can be found by quadrature.

The conditions to be satisfied by a $2n$-parameter family of curves in a space of $n+1$ dimensions if it is to represent the extremals of a problem with a prescribed transversality are at once apparent. The n expressions (18) for g_x/g must be identical and, likewise, the expressions (19) for $(1/g)\,\partial g/\partial y_k$. Moreover, these must actually be independent of the constants, that is,

$$\frac{\partial}{\partial \alpha_i}\left(\frac{1}{\partial m_k/\partial a_j}\,\frac{\partial}{\partial a_j}\Phi_k\right) \equiv 0 \qquad (i, j = 1, 2, \cdots, n). \tag{20}$$

Also, if the expressions (18) and (19) are actually to represent a function we must have

$$\begin{aligned} \frac{\partial^2 \log g}{\partial y_k \partial x} &= \frac{\partial^2 \log g}{\partial x \partial y_k}, \\ \frac{\partial^2 \log g}{\partial y_k \partial y_h} &= \frac{\partial^2 \log g}{\partial y_h \partial y_k}. \end{aligned} \tag{21}$$

Finally, the slope functions of the given curves must satisfy $w_{y_h y_k} = w_{y_k y_h}$, or

$$\frac{\partial}{\partial y_k}\left(\frac{m_k g e^{\Theta}}{1 + \sum m_i p_i}\right) = \frac{\partial}{\partial y_k}\left(\frac{m_k g e^{\Theta}}{1 - \sum m_i p_i}\right). \tag{17'}$$

These conditions might have been imposed along with (17). They were omitted at that time because, as we shall now show, they are satisfied by the same function, g, if they are satisfied at all.

Let us suppose that the equations (17) are satisfied by $g(x, y_1, y_2, \cdots, y_n)$. If we multiply the kth and the hth equations by m_h and $-m_k$, respectively, and add, the result is

$$\begin{aligned} &m_h \frac{\partial}{\partial y_k}\left(\frac{g e^{\Theta}}{1 + \sum m_i p_i}\right) - m_k \frac{\partial}{\partial y_h}\left(\frac{g e^{\Theta}}{1 + \sum m_i p_i}\right) \\ &= \frac{\partial e^{\Theta}}{1 + \sum m_i p_i}\left(m_h \frac{\partial m_k}{\partial x} - m_k \frac{\partial m_h}{\partial x}\right) = \frac{g e^{\Theta}}{1 + \sum m_i p_i}\left(\frac{w_{y_h}}{w_x^2} w_{y_k x} - \frac{w_{y_k}}{w_x^2} w_{y_h x}\right). \end{aligned}$$

Now we make use of the fact that $w_{y_k y_h} - w_{y_h y_k} = 0$, by adding to the right member of the previous equation

$$\frac{g e^{\Theta}(w_{y_k y_h} - w_{y_h y_k})}{1 + \sum m_i p_i w_x}.$$

That equation then becomes

$$\begin{aligned} m_h \frac{\partial}{\partial y_k}\left(\frac{g e^{\Theta}}{1 + \sum m_i p_i}\right) - m_k \frac{\partial}{\partial y_h}\left(\frac{g e^{\Theta}}{1 + \sum m_i p_i}\right) &= \left(\frac{g e^{\Theta}}{1 + \sum m_i p_i}\right)\left[\frac{\partial}{\partial y_h}\left(\frac{w_{y_k}}{w_x}\right)\right. \\ \left. - \frac{\partial}{\partial y_k}\left(\frac{w_{y_h}}{w_x}\right)\right] &= \frac{g e^{\Theta}}{1 + \sum m_i p_i}\left(\frac{\partial}{\partial y_h} - m_k \frac{\partial m_h}{\partial y_k}\right). \end{aligned}$$

This final form is identical with (17′) and shows that these equations are satisfied by the same function, $g(x, y_1, y_2, \cdots, y_n)$, that satisfies (17).

5. We shall conclude this section by showing directly that if the conditions which we have enumerated are satisfied and an integrand function, $f = ge^{\Theta}$, has been determined, then the given curves satisfy the Euler equations and are the extremals of the problem.

We know that the slope functions of a particular sub-family of the given curves satisfy*

$$\frac{\partial}{\partial y_i}[f - \sum p_k f_{p_k}] = \frac{\partial}{\partial x}[f_{p_i}], \tag{17}$$

$$\frac{\partial}{\partial y_i}[f_{p_k} = \frac{\partial}{\partial y_k}[f_{p_i}]. \tag{17'}$$

When we carry out the differentiation (17) becomes

$$f_{y_i} + \sum f_{p_k} p_{ky_i} - \sum p_{ky_i} f_{p_k} - \sum p_k \frac{\partial}{\partial y_i}[f_{p_k}] - f_{p_i x} - \sum]_{p_i p_k} p_{kx} = 0.$$

Then, by omitting terms which cancel and making use of (17′) we find

$$f_{y_i} - f_{p_i x} - \sum \frac{\partial}{\partial y_k}[f_{p_i}] p_k - \sum f_{p_i p_k} p_{kx} = 0,$$

which may be written

$$f_{y_i} - \frac{d}{dx} f_{p_i} = 0 \qquad (i = 1, 2, \cdots, n).$$

Since these are the Euler equations of the problem the given curves are actually the extremals of the integrand function which we have determined by means of them.

Example IV. Let us suppose that the set of functions which determine the transversality are

$$m_i = \frac{2p_i}{x^2 - \sum p_k^2} \qquad (i = 1, 2, \cdots, n).$$

First, we show that this set satisfies equations (5b) and is really a transversality. The left member of that equation becomes

$$\sum p_k \left[\frac{2p_k}{x^2 - \sum p_k^2} \frac{4p_i p_j}{(x^2 - \sum p_k^2)^2} - \frac{2p_i}{(x^2 - \sum p_k^2)} \frac{4p_j p_k}{(x^2 - \sum p_k^2)^2} \right] + \frac{4p_i p_j}{x^2 - \sum p_k^2}.$$

* In this paragraph the use of the square brackets, [ ], indicates that the differentiation is to be carried out after p has been replaced by $p(x, y_1, y_2, \cdots, y_n, \alpha_1, \alpha_2, \cdots, \alpha_n)$.

The terms under the summation sign are all zero and the last term is symmetrical in $p_i p_j$. Since substitution in the right member would give the same result the condition is satisfied.

Now we may calculate Θ by means of (9) and find

$$\Theta = \log (x^2 + \sum p_i^2).$$

Equation (8) then gives us the integrand function, which is

$$f = g(x, y)(x^2 + \sum p_i^2).$$

We shall now assume, further, that the extremals are given by the equations

$$y_i = (a_i x^3 - c_i)/3 \qquad (i = 1, 2, \cdots, n).$$

We can form a field by giving fixed values to the a_i. In this case the slope functions are

$$p_i = a_i x^2.$$

Then we calculate Φ_k by means of (17). This gives, if we denote $\sum a_i^2$ by A^2,

$$\Phi_k = -\frac{4a_k x A^2}{1 - x^2A^2} - \frac{2a_k}{1 - x^2A^2}\left(\frac{2}{x} + \frac{2xA^2}{1 + x^2A^2} - \frac{2xA}{1 + x^2A^2} - \frac{2xA^2}{1 - x^2A^2}\right)$$

$$= -\frac{2a_k}{1 - x^2A^2}\,\frac{2}{x} = -m_k\frac{2}{x}.$$

Substituting these results in (18) and (19) we find

$$\frac{g_x}{g} = -\frac{2}{x}, \text{ and } \frac{g_{y_k}}{g} = 0.$$

Omitting the constant of integration we have, therefore,

$$g = 1/x^2,$$

and

$$f = (1/x)^2(x^2 + \sum p_i^2).$$

The conditions which we have found are all satisfied and the given curves are actually the extremals of the integrand function.*

* Since this paper was prepared, Douglas has published a paper in these Transactions (April, 1927), entitled *Extremals and transversality of the general calculus of variations problem of the first order in space.* The subject matter of his paper has a good deal in common with that of the present paper, but the problem is given a geometric rather than an analytic interpretation.

YALE UNIVERSITY,
NEW HAVEN, CONN.

THE PROJECTIVE DIFFERENTIAL GEOMETRY OF SYSTEMS OF LINEAR HOMOGENEOUS DIFFERENTIAL EQUATIONS OF THE FIRST ORDER*

BY
ERNEST P. LANE

I. INTRODUCTION

The tedious calculations that occasionally occur in studying projective differential geometry by means of the invariants and covariants of a completely integrable system of linear homogeneous differential equations are admittedly an unpleasant feature of this method. But it seems that labor can sometimes be avoided by reducing the system of equations to a system of the first order. Such a system proves to be especially suitable for the study of a configuration composed of a set of varieties generated by linear spaces, with the generators in correspondence.

The present paper is concerned for the most part with the projective differential geometry that can be studied by means of a system of $n+1$ linear homogeneous differential equations of the first order in $n+1$ dependent variables and one independent variable. The transformation of dependent variables that is used is determined by the configuration to be studied, which in any case consists of a set of varieties each of which is generated by ∞^1 linear spaces, with the generators in correspondence. The precise number of possible configurations of this type in a space of a given number of dimensions is determined, but by imposing the condition that some of the varieties shall be covariant to the rest of them the range of applicability of the method can be widely extended. Certain curves called *intersector curves* on a variety are defined, and the locus of their tangents is investigated.

A section is devoted to geometry in a space S_5 of five dimensions. The intersection of a configuration with a hyperquadric in S_5 yields a geometric interpretation in ordinary ruled space. Another section is taken up with a pair of ruled surfaces in ordinary space, with their generators in correspondence. A complete system of invariants is furnished therefor, and a new canonical form of the differential equations is established.

Finally a system of $k(n+1)$ linear homogeneous partial differential equations of the first order in $n+1$ dependent variables and k independent var-

* Presented to the Society, December 29, 1927; received by the editors in December, 1927.

iables is considered, and the foundations are laid for the projective differential geometry of a configuration composed of a set of varieties each of which is generated by ∞^k linear spaces, with the generators in correspondence.

II. Ordinary differential equations

In a linear space S_n of n dimensions let us consider $n+1$ linearly independent points P_i with projective homogeneous coördinates

$$x_i^{(1)}, \cdots, x_i^{(n+1)} \qquad (i = 1, \cdots, n+1),$$

which are single-valued analytic functions of one independent variable t. As t varies each point P_i describes a curve C_i, and we thus obtain $n+1$ curves in S_n with their points in correspondence, corresponding points being those that correspond to the same value of the parameter t.

The square matrix of the coördinates $x_i^{(j)}$ of the points P_i is of order and rank $n+1$. Therefore it is possible to determine the coefficients of a system of equations of the form

$$(E_n) \qquad x_i' = \sum_{j=1}^{n+1} a_{ij}x_j \qquad (i = 1, \cdots, n+1),$$

so that $(x_1^{(k)}, \cdots, x_{n+1}^{(k)})$ $(k=1, \cdots, n+1)$ will be $n+1$ sets of solutions. For example we may substitute each of these sets in turn in the first equation of system (E_n) and then solve the resulting $n+1$ equations for the coefficients of the first of equations (E_n). Similarly, the coefficients of each of the other n equations can be determined.

The transformation of dependent variables

$$(T_{n\cdot 1}) \qquad x_i = \lambda_i \bar{x}_i \qquad (i = 1, \cdots, n+1),$$

and the transformation of independent variable

$$(U) \qquad u = u(t), \quad u' \neq 0,$$

leave each of the curves C_i invariant. Therefore *the projective differential geometry of a set of $n+1$ curves in S_n with their points in correspondence can be studied by means of the invariants and covariants of system (E_n) under the total transformation $(T_{n\cdot 1}U)$.* Various configurations covariant to such curves can be defined at once. For example, n of the points P_i corresponding to a value of t determine a hyperplane S_{n-1}, and as t varies this hyperplane osculates a curve. In this way $n+1$ new curves with their points in correspondence with those of the original curves are obtained. And *for each k $(k=1, \cdots, n-1)$ there are $C_{n+1,k+1}$ covariant varieties V_{k+1} each of which is generated by ∞^1 linear spaces S_k determined by $k+1$ of the points P_i, and the*

generators of all of these are in correspondence with the points P_i of the curves C_i.

With a different choice of the transformation of dependent variables, the projective differential geometry of a different configuration can be studied by means of the corresponding invariants and covariants of system (E_n). The situation will be made clear by illustrations. If $n=2$, besides the geometry of a triple of plane curves* with their points in correspondence, we can also study by means of (E_2) the geometry of *a pair of plane curves with the tangents of one corresponding to the points of the other* if we use the transformation of dependent variables

$$(T_{2\cdot 2}) \qquad x_i = \sum_{j=1}^{2} \lambda_{ij}\bar{x}_j \quad (i = 1,2), \quad x_3 = \lambda_3\bar{x}_3.$$

Obviously the points of these two curves are in correspondence, and the lines joining corresponding points envelop a new covariant curve, while the point of intersection of the tangents at corresponding points generates another, and so on.

If $n=3$, there are three possibilities besides a quadruple of space curves with their points in correspondence. First of all, if we use the transformation

$$(T_{3\cdot 2}) \qquad x_i = \sum_{j=1}^{2} \lambda_{ij}\bar{x}_j, \quad x_k = \lambda_k\bar{x}_k \qquad (i = 1,2\,;\ k = 3,4),$$

we obtain the geometry of *a ruled surface and a pair of curves with the generators of the surface and the points of the curves in correspondence.* The lines joining corresponding points of the curves generate a covariant ruled surface. In the second place, the transformation

$$(T_{3\cdot 3}) \qquad x_i = \sum_{j=1}^{2} \lambda_{ij}\bar{x}_j, \quad x_k = \sum_{j=3}^{4} \lambda_{kj}\bar{x}_j \qquad (i = 1,2\,;\ k = 3,4)$$

yields the geometry of *a pair of ruled surfaces with generators in correspondence* which we have studied elsewhere,† and concerning which we shall have more to say in §4. Finally, the transformation

$$(T_{3\cdot 4}) \qquad x_i = \sum_{j=1}^{3} \lambda_{ij}\bar{x}_j \ (i = 1,2,3), \quad x_4 = \lambda_4\bar{x}_4$$

gives *a developable surface and a curve with the tangent planes of the one and the points of the other in correspondence.* The points of the edge of regression

* Jerbert, Abstract, Bulletin of the American Mathematical Society, vol. 33 (1927), p. 516.

† Lane, *Ruled surfaces with generators in one-to-one correspondence*, these Transactions, vol. 25 (1923), p. 281.

of the developable and those of the other curve are in correspondence, and the ruled surface of lines joining corresponding points is a covariant ruled surface, as is also the ruled surface of lines of intersection of the osculating planes of the two curves at corresponding points.

In general, *the number of configurations whose projective differential geometry can be studied in S_n by means of the invariants and covariants of system (E_n) under linear transformations of the dependent variables and the transformation (U) of the independent variable is $n(n^2+20)/24$ if n is even and $n \geqq 2$; and the number is $n(n^2+23)/24$ if n is odd.* These formulas can be obtained by observing that the number under consideration is the number of ways in which it is possible to choose at least two transformations of the form

$$(1)\qquad x_i = \sum_{j=1}^{a} \lambda_{ij}\bar{x}_j \ (i = 1, \cdots, a)\ ;\ x_k = \sum_{j=a+1}^{b} \lambda_{kj}\bar{x}_j \quad (k = a+1, \cdots, b)\ ;$$
$$\cdots\ ;\ x_l = \sum_{j=g+1}^{h} \lambda_{lj}\bar{x}_j \quad (l = g+1, \cdots, h),$$

where $a, b, \cdots, g, h$ are positive integers such that

$$a \geqq b - a \geqq \cdots \geqq g - f \geqq h - g, \quad a < b < \cdots < g < h = n+1.$$

In particular, if $n=4$, the number of configurations is 6, and if $n=5$, the number is 10.

Let us consider any one of the possible configurations in S_n, and any one of the varieties V_{k+1} in this configuration which is the locus of an S_k with $1 \leqq k \leqq n-2$. A curve on such a variety will be called an *intersector curve* with respect to the remaining varieties in the configuration in case the tangent to the curve at the point where it crosses a generator S_k intersects the linear space S_{n-k-1} determined by the generators of the other varieties in the configuration that correspond to S_k. In order to obtain the differential equations of these curves let us observe that any curve on the V_{k+1} can be regarded as the locus of a point ϕ, where

$$(2)\qquad \phi = x_1 + \sum_{j=1}^{k+1} \lambda_j x_j.$$

The point $\phi' + h\phi$ is any point on the tangent of this curve at the point ϕ, and $\phi' + h\phi$ can be expressed as a linear combination of $x_1, \cdots, x_{n+1}$. If the curve is an intersector curve, then the coefficients of $x_1, \cdots, x_{k+1}$ must vanish. Thus we obtain $k+1$ equations, and elimination of h from them gives *the differential equations of the intersector curves on V_{k+1}*

$$(3)\quad \lambda_p' + a_{1p} - a_{11}\lambda_p + \sum_{j=2}^{k+1} \lambda_j(a_{jp} - \lambda_p a_{j1}) = 0 \qquad (p = 2, \cdots, k+1).$$

There are ∞^k of these curves, one of them passing through each point of a generator S_k.

Tangents of intersector curves may be called *intersector tangents*. The locus of the intersector tangents at points of a fixed generator S_k is a variety V_{k+1}, whose equations can be found in the following way. In the expression for $\phi' + h\phi$ as a linear combination of $x_1, \cdots, x_{n+1}$ let us replace λ_p' by the expression given therefor by equation (3). Then let us denote the coefficients of $x_1, \cdots, x_{n+1}$ by $y_1, \cdots, y_{n+1}$ respectively, so that $y_1, \cdots, y_{n+1}$ are the local coördinates of the point $\phi' + h\phi$ on an intersector tangent referred to the pyramid whose vertices are the points $x_1, \cdots, x_{n+1}$. Thus we find *the parametric equations of the locus of the intersector tangents at points of a fixed generator,*

$$(4)\quad \begin{aligned} y_1 &= a_{11} + \sum_{j=2}^{k+1} \lambda_j a_{j1} + h, \quad y_p = \lambda_p y_1 \quad (p = 2, \cdots, k+1), \\ y_q &= a_{1q} + \sum_{j=2}^{k+1} \lambda_j a_{jq} \qquad\qquad (q = k+2, \cdots, n+1). \end{aligned}$$

If $2k > n-2$, homogeneous elimination of $\lambda_2, \cdots, \lambda_{k+1}$ and h from these equations gives $n-k-1$ quadratic equations of the locus of the intersector tangents,

$$(5)\quad \sum_{j=1}^{k+1} y_j(a_{jq}y_{n+1} - y_q a_{j,n+1}) = 0 \quad (q = k+2, \cdots, n).$$

The locus is therefore a V_{k+1} of order $2(n-k-1)$. If, however, $2k \leqq n-2$, homogeneous elimination gives $n-1-2k$ linear equations and k quadratic equations. In this case the locus is a V_{k+1} of order $2k$.

III. Space of five dimensions

The case $n=5$ offers especial interest because the intersection of a configuration with a hyperquadric in S_5 can be interpreted as a configuration of line geometry in ordinary ruled space. And of the ten possible configurations in S_5 there are three which, because of their symmetrical properties, seem to be most interesting.

Denoting the equations (E_n) when $n=5$ by (E_5), let us consider in connection therewith the transformation

$$(T_{5\cdot 4})\quad x_i = \sum_{j=1}^{2} \lambda_{ij}\bar{x}_j, \quad x_k = \sum_{j=3}^{4} \lambda_{kj}\bar{x}_j, \quad x_l = \sum_{j=5}^{6} \lambda_{lj}\bar{x}_j$$
$$(i = 1, 2; \quad k = 3, 4; \quad l = 5, 6).$$

The configuration to be studied is *a triple of ruled surfaces with their generators in correspondence.* One of these ruled surfaces R_{12} is the locus of the line l_{12} joining the points P_1, P_2; and similar notations will be used for the others.

Considering a point $x_1+\lambda x_2$ on l_{12} we find that the differential equation of the intersector curves on R_{12} with respect to R_{34} and R_{56} is

$$\lambda' + a_{12} - (a_{11} - a_{22})\lambda - a_{21}\lambda^2 = 0. \tag{6}$$

The harmonic property of four particular solutions of an equation of Riccati can be given a new geometric interpretation here. And if λ is the general solution of (6) the intersector tangents of R_{12} at points of l_{12} intersect the S_3 determined by the corresponding generators l_{34}, l_{56} in a straight line which is the locus of the point

$$\sum_{j=3}^{6} (a_{1j} + \lambda a_{2j})x_j$$

as the constant of integration takes all possible values. As t varies, this line generates a ruled surface on the V_4 generated by the S_3.

The osculating plane of a curve on R_{12} generated by a point $x_1+\lambda x_2$ on l_{12} ordinarily meets the S_3 of the corresponding generators l_{34}, l_{56} in a point whose coördinates can easily be found. But if we impose on the curve the condition that the intersection of its osculating plane and the S_3 shall be a straight line, then there are only two curves of this type on R_{12}, and for them λ is a root of the equation

$$\sum_{j=3}^{6} [a_{2j}a_{j1}\lambda^2 + (a_{1j}a_{j1} - a_{2j}a_{j2})\lambda - a_{1j}a_{j2}] = 0. \tag{7}$$

The effect of $(T_{5\cdot 4})$ on system (E_5) is to produce another system of the same form. If we choose $(\lambda_{11}, \lambda_{21})$ and $(\lambda_{12}, \lambda_{22})$ as two pairs of solutions of the differential equations

$$\phi' = a_{11}\phi + a_{12}\psi, \qquad \psi' = a_{21}\phi + a_{22}\psi, \tag{8}$$

we find that $\bar{a}_{ij}=0$ $(i, j=1, 2)$. Similarly we can make $\bar{a}_{ij}=0$ $(i, j=3, 4)$, and $\bar{a}_{ij}=0$ $(i, j=5, 6)$. Thus we obtain a canonical form of system (E_5) for which the directrix curves on each surface are intersector curves with respect to the other two, and for which, moreover, the derivative point of a point x_i on one surface lies in the S_3 determined by the generators of the other two surfaces that correspond to the generator through x_i.

Another canonical form of (E_5) can be obtained in the following way. Let us define the symbol (xy) by the equation

$$(xy) = x^{(1)}y^{(6)} + x^{(2)}y^{(5)} + x^{(3)}y^{(4)} + x^{(4)}y^{(3)} + x^{(5)}y^{(2)} + x^{(6)}y^{(1)}. \tag{9}$$

Then $(xx)=0$ is the equation of a hyperquadric V_4^2 in S_5, and a point $\lambda x_1+\mu x_2$ on l_{12} lies on V_4^2 in case $\lambda:\mu$ satisfies the equation

$$(x_1x_1)\lambda^2 + 2(x_1x_2)\lambda\mu + (x_2x_2)\mu^2 = 0. \tag{10}$$

Assuming that the roots of this equation are distinct, so that l_{12} is not tangent to V_4^2, let us choose one root for $\lambda_{11}:\lambda_{12}$ and the other for $\lambda_{21}:\lambda_{22}$. Then the transformation $(T_{5\cdot 4})$ makes the new directrix curves C_1, C_2 of R_{12} lie on V_4^2. In this way we can obtain a canonical form of (E_5) for which all six directrix curves lie on V_4^2.

The usefulness of the second canonical form is due to the well known fact that point geometry on V_4^2 in S_5 is isomorphic with line geometry in ordinary ruled space R_3. To the curves C_1, C_2 on V_4^2 correspond two ruled surfaces in R_3. *The geometry of the intersection of V_4^2 and a triple of ruled surfaces in S_5 with their generators in correspondence is equivalent to the geometry of a triple of pairs of ruled surfaces in R_3 with their generators in correspondence.* The assumption that the points P_i are linearly independent implies that six corresponding generators in R_3 do not belong to a linear complex.

A triple of ruled surfaces* in R_3 with their generators in correspondence is equivalent to a triple of curves on V_4^2 in S_5 with their points in correspondence. If the ruled surfaces of the triple are given by their parametric equations in line coördinates

$$x_i^{(j)} = x_i^{(j)}(t) \quad (i = 1,2,3\,;\ j = 1, \cdots, 6),\ (x_ix_i) = 0, \tag{11}$$

and if no three corresponding generators with the generators consecutive to each of them belong to the same linear complex, then on V_4^2 we have three curves whose tangents at corresponding points do not lie in an S_4. Therefore, the coördinates $(x_1^{(j)}, x_2^{(j)}, x_3^{(j)})$ $(j=1, \cdots, 6)$ form six sets of solutions of a system of differential equations of the form

$$x_i'' = \sum_{j=1}^{3} (p_{ij}x_j' + q_{ij}x_j) \qquad (i = 1,2,3), \tag{12}$$

and the geometry of the triple of ruled surfaces can be studied by means of the invariants and covariants of these equations under the group of transformations

$$x_i = \lambda_i\bar{x}_i \quad (i = 1,2,3), \quad u = u(t). \tag{13}$$

Equations (12) can be converted into a system of the form (E_5) by placing $x_4=x_1'$, $x_5=x_2'$, $x_6=x_3'$. The appropriate group of transformations is readily determined.

* Carpenter, *Triads of ruled surfaces*, these Transactions, vol. 29 (1927), p. 254.

Let us now consider system (E_5) and the transformation

$$(T_{5\cdot7}) \qquad x_i = \sum_{j=1}^{3} \lambda_{ij}\bar{x}_j, \quad x_k = \sum_{j=4}^{6} \lambda_{kj}\bar{x}_j \quad (i = 1,2,3\,;\, k = 4,5,6).$$

The configuration in this case is *a pair of varieties V_3, each of which is generated by ∞^1 planes, with the generators in correspondence.* We shall denote these varieties by V_3' and V_3'', the first being generated by the plane S_2' determined by the points P_1, P_2, P_3, and the second by S_2'' determined by P_4, P_5, P_6.

By an easy extension of the method used in connection with equations (8) it is possible to obtain a canonical form of system (E_5) for which $\bar{a}_{ij} = \bar{a}_{pq} = 0$ $(i, j = 1, 2, 3;\ p, q = 4, 5, 6)$. The directrix curves on each variety are then intersector curves with respect to the other, and the derivative point of a point x_i on one variety lies in the corresponding generating plane of the other.

Another canonical form of system (E_5) can be obtained by demanding that the directrix curves of V_3' and V_3'' lie on the hyperquadric V_4^2. A point $\lambda_1x_1+\lambda_2x_2+\lambda_3x_3$ on S_2' lies on the conic of intersection of S_2^2 and V_4^2 in case $\lambda_1, \lambda_2, \lambda_3$ satisfy the equation

$$(14) \qquad \sum_{i,j=1}^{3} (x_ix_j)\lambda_i\lambda_j = 0.$$

Therefore three sets of values of $\lambda_1, \lambda_2, \lambda_3$ which satisfy this equation furnish three directrix curves of V_3' on V_4^2. Similarly, three directrix curves for V_3'' can be found on V_4^2.

To the conic of intersection of S_2' and V_4^2 corresponds in R_3 a regulus composed of one family of generators of a quadric surface. As t varies, the conic generates a surface on V_4^2, and the regulus generates a congruence. The geometry of the intersection of the hyperquadric and the two varieties V_3', V_3'' is equivalent to the geometry of two congruences each of which is the locus of ∞^1 reguli in R_3, with generating reguli in correspondence. Ordinarily two consecutive reguli of such a congruence do not intersect. But the flecnode congruence of a ruled surface is an example of such a congruence in which two consecutive reguli do intersect in a generator of the ruled surface counted twice. And the congruence of asymptotic tangents of a ruled surface is an example of such a congruence in which two consecutive reguli intersect in two flecnode tangents of the ruled surface.

If we consider finally system (E_6) and the transformation

$$(T_{6\cdot9}) \qquad x_i = \sum_{j=1}^{4} \lambda_{ij}\bar{x}_j, \quad x_k = \sum_{j=5}^{6} \lambda_{kj}\bar{x}_j \quad (i = 1, \cdots, 4\,;\, k = 5,6),$$

the configuration is *a V_4, which is the locus of ∞^1 linear spaces S_3, and a ruled surface R_{56}, with generators in correspondence.*

A canonical form of system (E_5) can be obtained for which $\bar{a}_{ij}=\bar{a}_{pq}=0$ $(i, j=1, \cdots, 4;\ p, q=5, 6)$. This canonical form is related to the intersector curves on V_4 and R_{56}, as in previous cases.

The intersection of this configuration with V_4^3 is a variety V_3 generated by ∞^1 quadrics V_2^2, and two curves which are the loci of points in correspondence with the quadrics. The geometry of this configuration on the hyperquadric is equivalent in R_3 to the geometry of a complex generated by ∞^1 linear congruences, and a pair of ruled surfaces whose generators are in correspondence with the congruences. Two consecutive congruences of the complex ordinarily intersect in two lines which generate two ruled surfaces in the complex. Moreover, the axes of a congruence of the complex also generate two ruled surfaces, and these are such that their two common tangents at points of corresponding generators are the aforesaid lines of intersection of two consecutive congruences.

An example of a complex of the type under consideration is the complex generated by the osculating linear congruences of a ruled surface. Such a congruence is determined by four consecutive generators of the ruled surface, and consists of the lines intersecting two flecnode tangents. Two consecutive congruences intersect in an osculating regulus determined by three consecutive generators of the ruled surface. Three consecutive congruences intersect in a generator of the surface.

A second example is the complex of tangents of a ruled surface. This consists of ∞^1 special linear congruences of tangents at points of a generator. Two consecutive congruences of this kind intersect in a regulus of asymptotic tangents, and three consecutive congruences intersect in two flecnode tangents at points of a generator of the ruled surface.

IV. Pairs of ruled surfaces in ordinary space

In the paper referred to in §2 we studied a pair of ruled surfaces in ordinary space with their generators in correspondence, using a transformation of the form $(T_{3\cdot 3})$, and using the following differential equations:

E_3)
$$\begin{aligned} y' &= c_{11}y + c_{12}z + a_{11}\rho + a_{12}\sigma, \\ z' &= c_{21}y + c_{22}z + a_{21}\rho + a_{22}\sigma, \\ \rho' &= b_{11}y + b_{12}z + d_{11}\rho + d_{12}\sigma, \\ \sigma' &= b_{21}y + b_{22}z + d_{21}\rho + d_{22}\sigma. \end{aligned}$$

For every configuration studied by our method the problem of computing

a complete system of invariants arises. The method ordinarily used is that of infinitesimal transformations according to Lie's theory of continuous groups. This method applied to system (E_3) leads to the following results. *A complete system of seminvariants*, i.e., invariants under $(T_{3\cdot 3})$, *of a pair of ruled surfaces with generators in correspondence consists of the eight seminvariants* $I_2, I_4, J_4, J_8, K_3, K_5, K_7, L_6$ *defined by the following formulas*:

$$I_2 = \sum a_{ij}b_{ij},\ J_4 = \sum A_{ij}B_{ij},\ K_3 = \sum A_{ij}b_{ij},\ L_6 \equiv \sum \mathfrak{A}_{ij}\mathfrak{B}_{ij} \qquad (i,j = 1,2),$$

$$I_4 = (a_{11}a_{22} - a_{12}a_{21})(b_{11}b_{22} - b_{12}b_{21}),$$

$$J_8 = (A_{11}A_{22} - A_{12}A_{21})(B_{11}B_{22} - B_{12}B_{21}), \tag{15}$$

$$K_5 = (a_{11}a_{22} - a_{12}a_{21})(B_{11}b_{22} - B_{12}b_{21} - B_{21}b_{12} + B_{22}b_{11}),$$

$$K_7 = (A_{11}A_{22} - A_{12}A_{21})(B_{11}b_{22} - B_{12}b_{21} - B_{21}b_{12} + B_{22}b_{11}),$$

where

$$\begin{aligned} A_{11} &= a'_{11} - a_{11}c_{11} - a_{21}c_{12} + a_{11}d_{11} + a_{12}d_{21}, \\ A_{12} &= a'_{12} - a_{12}c_{11} - a_{22}c_{12} + a_{11}d_{12} + a_{12}d_{22}, \\ A_{21} &= a'_{21} - a_{11}c_{21} - a_{21}c_{22} + a_{21}d_{11} + a_{22}d_{21}, \\ A_{22} &= a'_{22} - a_{12}c_{21} - a_{22}c_{22} + a_{21}d_{12} + a_{22}d_{22}, \end{aligned} \tag{16}$$

and the formulas for B_{ij} are obtained from those for A_{ij} by interchanging a_{ij} and b_{ij}, c_{ij} and d_{ij}, while the formulas for $\mathfrak{A}_{ij}$ and $\mathfrak{B}_{ij}$ are obtained from those for A_{ij} and B_{ij} respectively by replacing therein a_{ij} by A_{ij}, and b_{ij} by B_{ij}. Every seminvariant is a function of these and their derivatives.

A complete system of invariants consists of the eight invariants $I_2, I_4, \theta_3, \theta_5, \theta_6, \theta_{10}, \phi_{10}, \theta_8$, *of which the last six are defined by the following formulas*:

$$\begin{aligned} \theta_3 &= I_2' - 2K_3,\ \theta_5 = I_4' - 2K_5,\ \theta_6 = 4I_2J_4 - I_2'^2, \\ \theta_{10} &= 3K_5^2 - 2K_5I_4' + 4I_4^2K_7/K_5, \\ \phi_{10} &= 4I_4K_5J_8/K_7 - K_5^2, \\ \theta_8 &= I_2''(I_2'' - 4J_4) - 4I_2L_6 - 2J_4^2 - 3I_2'^2I_2''/I_2 \\ &\quad + 6I_2'J_4' + 15I_2'^4/8I_2^2. \end{aligned} \tag{17}$$

All of the invariants can be calculated by application of the jacobian process* to these eight fundamental invariants.

By means of $(T_{3\cdot 3})$ it is possible to reduce system (E_3) to the canonical form

$$\begin{aligned} y' &= c_{12}z + a_{11}\rho, \qquad \rho' = b_{11}y + d_{12}\sigma, \\ z' &= c_{21}y + a_{22}\rho, \qquad \sigma' = b_{22}z + d_{21}\rho. \end{aligned} \tag{18}$$

* Wilczynski, *Projective Differential Geometry of Curves and Ruled Surfaces*, p. 23.

For this canonical form the tetrahedron $yz\rho\sigma$ is such that the two faces through l_{yz} are tangent to R_{yz} at P_y and P_z, while the two faces through $l_{\rho\sigma}$ are tangent to $R_{\rho\sigma}$ at P_ρ and P_σ. The lines $l_{y\rho}$ and $l_{z\sigma}$ are the common tangents of R_{yz} and $R_{\rho\sigma}$. The two derivative points z' and ρ' are on $l_{y\sigma}$, while the points y' and σ' are on $l_{z\rho}$. The anharmonic ratios $(y\sigma z'\rho')$ and $(z\rho y'\sigma')$ are respectively equal to $a_{22}b_{11}/d_{12}c_{21}$ and $a_{11}b_{22}/c_{12}d_{21}$. A complete system of invariants for this canonical form consists of the eight invariants

$$a_{11}b_{11},\ a_{22}b_{22},\quad c_{12}c_{21},\ d_{12}d_{21},\ a_{11}d_{12}c_{21}/a_{22},$$

$$a'_{11}b_{11} - a_{11}b'_{11},\ a'_{22}b_{22} - a_{22}b'_{22},\ c'_{12}c_{21} - c_{12}c'_{21}.$$

It is not difficult to show that *if these invariants are given as functions of t, they determine a pair of ruled surfaces with generators in correspondence, except for a projective transformation.*

A ruled surface and a derivative* ruled surface are a pair of ruled surfaces with their generators in correspondence. The derivative of R_{yz} is generated by $l_{\rho\sigma}$, where

$$\rho = 2y' + p_{11}y + p_{12}z, \qquad \sigma = 2z' + p_{21}y + p_{22}z, \tag{19}$$

the coefficients p_{ij} being those of Wilczynski's system (A). It is easy to show that in this case the coefficients of system (E_3) are given by the formulas

$$c_{ij} = d_{ij} = -\tfrac{1}{2}p_{ij},\ \ b_{ij} = \tfrac{1}{2}u_{ij},\ \ a_{ij} = \tfrac{1}{2},\ a_{12} = a_{21} = 0 \quad (i,j = 1,2), \tag{20}$$

wherein the functions u_{ij} are those usually denoted† by these symbols. If for $R_{\rho\sigma}$ we choose the principal derivative ruled surface‡ of R_{yz}, then our theory of a pair of ruled surfaces becomes a theory of the single ruled surface R_{yz}, because the principal derivative of a ruled surface is covariant to the surface.

V. Partial differential equations

If the coördinates $x_i^{(j)}$ $(i,j=1,\cdots,n+1)$ of $n+1$ points P_i are functions of k independent variables $u^1,\cdots,u^k$ $(k<n)$, then the locus of each point is a variety V_k, and we thus obtain $n+1$ such varieties in S_n with their points in correspondence. It is possible to determine the coefficients of a system of equations of the form

$$\frac{\partial x_i}{\partial u^p} = \sum_{j=1}^{n+1} a_{ijp}x_j \qquad (i = 1,\cdots,n+1;\ p = 1,\cdots,k), \tag{21}$$

* Wilczynski, loc. cit., p. 146.

† Wilczynski, loc. cit., p. 96.

‡ Wilczynski, loc. cit., p. 216. Stouffer, *Some canonical forms and associated expansions in projective differential geometry*, Address before the Southwestern Section, St. Louis, Nov. 26, 1927.

so that $(x_1^{(j)}, \cdots, x_{n+1}^{(j)})$ $(j=1, \cdots, n+1)$ will be $n+1$ sets of solutions. These coefficients satisfy $\frac{1}{2}k(k-1)(n+1)^2$ integrability conditions which can be obtained by demanding that the derivatives of the second order calculated by differentiating equations (21) shall be unique. These integrability conditions are

$$\frac{\partial a_{ijp}}{\partial u^q} + \sum_{l=1}^{n+1} a_{ilp}a_{ljq} = \frac{\partial a_{ijq}}{\partial u^p} + \sum_{l=1}^{n+1} a_{ilq}a_{ljp} \tag{22}$$
$$(i,j = 1, \cdots, n+1\,;\ p,q = 1, \cdots, k\,;\ p < q).$$

From these conditions, it is easy to deduce the equations

$$\sum_{i=1}^{n+1}\left(\frac{\partial a_{iip}}{\partial u^q} - \frac{\partial a_{iiq}}{\partial u^p}\right) = 0 \qquad (p,q = 1, \cdots, k\,;\ p < q). \tag{23}$$

A transformation

$$x_i = \lambda_i \bar{x}_i \qquad (i = 1, \cdots, n+1), \tag{24}$$

with the λ_i supposed to be functions of $u^1, \cdots, u^k$, and a transformation

$$\bar{u}^p = \bar{u}^p(u^1, \cdots, u^k) \qquad (p = 1, \cdots, k), \tag{25}$$

leave each of the varieties V_k invariant. Therefore *the projective differential geometry of a set of $n+1$ varieties in S_n, with their points in correspondence, can be studied by means of the invariants and covariants of system* (22) *under the total transformation* (24), (25).

If we use a general transformation of the form (1) with coefficients that are functions of $u^1, \cdots, u^k$, we obtain the geometry of a configuration which consists of a certain number of varieties, each of which is the locus of ∞^k linear spaces in correspondence. For example, if $n=3$ and $k=2$, and if $(T_{3\cdot 1})$ is used we have the geometry of a quadruple of surfaces* in ordinary space, with their points in correspondence. If $n=3$ and $k=2$, but if $(T_{3\cdot 3})$ is used, the configuration is a pair of congruences† in ordinary space with their generators in correspondence.

---

* D. Sun, a student at Chicago, has elaborated the details of this geometry in a work as yet unpublished. This was his doctoral dissertation, August 31, 1928.

† A. J. Cook, a student at Chicago, is studying this configuration in his doctoral dissertation.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

THE BEHAVIOR OF A BOUNDARY VALUE PROBLEM AS THE INTERVAL BECOMES INFINITE*

BY
W. E. MILNE

The boundary value problems for linear self-adjoint differential equations of the second order with homogeneous linear boundary conditions at the ends of a finite interval have been extensively studied and the principal facts are well known. In this connection a problem of some interest arises if the ends of the interval at which the boundary conditions apply are allowed to recede to $-\infty$ and $+\infty$. The aim of this paper is to investigate the behavior of characteristic numbers, characteristic solutions, and oscillation properties, as the interval becomes infinite. A closely related problem for the differential equation

$$(d/dx)(p(x)du/dx) + (\lambda - q(x))u = 0$$

has been solved by Weyl† and further studied by Hilb‡ and Gray.§

In this paper some interesting results are obtained for the equation

$$d^2u/dx^2 + G(x, \lambda)u = 0$$

and are set forth in Theorems I and II.

It is planned to treat the degree of convergence of certain associated expansions for the infinite interval in a subsequent paper.

1. The differential equation under investigation is

$$d^2u/dx^2 + G(x,\lambda) = 0. \tag{1}$$

The function $G(x, \lambda)$ is assumed to be real and continuous and to possess a positive partial derivative with respect to λ for all real values of x and λ. It is further assumed that

$$\lim_{\lambda=-\infty} G(x,\lambda) = -\infty, \quad \lim_{\lambda=+\infty} G(x,\lambda) = +\infty, \tag{2}$$

and that

$$\lim_{x=\pm\infty} G(x,\lambda) = -\infty. \tag{3}$$

* Presented to the Society, San Francisco Section, June 2, 1928; received by the editors in December, 1927, and May, 1928.

† Mathematische Annalen, vol. 68 (1910), p. 220, and Göttinger Nachrichten, 1910, p. 442.

‡ Mathematische Annalen, vol. 76 (1915), p. 333.

§ American Journal of Mathematics, vol. 50 (1928), p. 431.

These conditions are all satisfied in the important special case

$$G(x,\lambda) = \lambda - q(x)$$

provided that $q(x)$ is real and continuous and $\lim q(x) = +\infty$.

Associated with (1) we consider boundary conditions at $x=a$ and $x=b$,

$$\begin{aligned} &\alpha_1 u(a) + \alpha_2 u'(a) + \alpha_3 u(b) + \alpha_4 u'(b) = 0, \\ &\beta_1 u(a) + \beta_2 u'(a) + \beta_3 u(b) + \beta_4 u'(b) = 0. \end{aligned} \tag{4}$$

It is assumed of course that these conditions are linearly independent. Let the determinant $\alpha_i\beta_j - \alpha_j\beta_i$ be denoted by A_{ij}, and let $A_{12} = A_{34}$, so that the conditions are self-adjoint. Then there exists* an infinite sequence of values of λ, $l_0, l_1, l_2, l_3, \cdots$, with limit point at $+\infty$ only, and furnishing solutions of (1) and (4). If we take these values in increasing order of magnitude, counting each double value twice, the solution $u_n(x)$ corresponding to $\lambda = l_n$ vanishes at least n times and not more than $n+2$ times in the interval $a < x \leqq b$.

Now as a and b recede to $-\infty$ and $+\infty$ respectively, what becomes of the l_n, the $u_n(x)$, and the roots of $u_n(x)$?

2. In order to answer the foregoing question we must recall some facts concerning the solutions of equation (1), their roots and their behavior at infinity. Multiply (1) by $u(x)$ and integrate from x_1 to x_2, integrating the first term by parts. After transposition we have

$$u(x_2)u'(x_2) - u(x_1)u'(x_1) = \int_{x_1}^{x_2} u'^2(x)dx - \int_{x_1}^{x_2} G(x,\lambda)u(x)dx. \tag{5}$$

From the hypothesis (3) it follows that there exist two numbers α and β such that $G(x, \lambda)$ is negative when $x > \beta$ and when $x < \alpha$. Then if $x_2 > x_1 > \beta$ the right hand side of (5) is positive, which shows that the product $u(x)u'(x)$ does not vanish at both x_1 and x_2. Stated otherwise, $u(x)u'(x)$ does not have more than one root greater than β and does not have more than one root less than α. The total number of roots of $u(x)$ is therefore finite since the interval between any two consecutive roots is not less than π/M, where M^2 is a constant such that $G(x, \lambda) \leqq M^2$.

Now it is quite important to show that as x becomes infinite†

$$\lim u'(x)/u(x) = \infty. \tag{6}$$

---

* See for example Birkhoff, these Transactions, vol. 10 (1909), p. 264.

† For the behavior of solutions at infinity see Wiman, Arkiv för Matematik, Astronomi och Fysik, vol. 12, No. 14.

To see this we note first of all that when x is greater than the greatest root of $u(x)$ the function $R=u'(x)/u(x)$ is continuous. Let N be any positive number, as large as we please, and let x be chosen so large that $-G(x, \lambda)$ remains greater than $2N^2$. By differentiation and substitution from (1) we get

$$dR/dx = -G(x,\lambda) - R^2,$$

so that if R ever comes within the interval $-N<R<N$, the derivative dR/dx will be greater than N^2. Consequently R will increase beyond N, and will not return, since R is continuous and dR/dx is positive for $R=N$. Thus (6) is established.

Let the principal solutions of equation (1) at the origin be denoted by $u_1(x)$ and $u_2(x)$ so that

$$u_1(0) = u_2'(0) = 1, \quad u_2(0) = u_1'(0) = 0. \tag{7}$$

These solutions satisfy the well known identity

$$u_1(x)u_2'(x) - u_2(x)u_1'(x) = 1. \tag{8}$$

Then the general solution of equation (1) can be written

$$u(x) = C(u_1^2 + u_2^2)^{1/2} \sin [\phi(x) - \theta], \tag{9}$$

in which C and θ are arbitrary constants and $\phi(x)$ is defined by the equation

$$\phi(x) = \tan^{-1} u_2(x)/u_1(x). \tag{10}$$

By differentiating (10) and using (8) we get

$$d\phi/dx = [u_1^2 + u_2^2]^{-1},$$

so that $d\phi/dx$ is positive and ϕ is an increasing function of x. In view of the fact that $u(x)$ has a finite number of roots, $\phi(x)$ cannot increase indefinitely, and therefore must approach a limit. The same conclusion applies as x becomes negatively infinite. We therefore may define ϕ_1 and ϕ_2 as follows:

$$\phi_1 = \lim_{x=-\infty} \phi(x), \quad \phi_2 = \lim_{x=\infty} \phi(x).$$

It is now convenient to define two pairs of independent particular solutions of equation (1) as follows:

$$\begin{aligned} V_1(x) &= (u_1^2 + u_2^2)^{1/2} \sin [\phi(x) - \phi_1], \\ W_1(x) &= (u_1^2 + u_2^2)^{1/2} \cos [\phi(x) - \phi_1], \end{aligned} \tag{11}$$

and another pair V_2 and W_2 similarly defined with ϕ_2 in place of ϕ_1. As x approaches $-\infty$ it can be shown that

(12) $$\lim V_1(x) = \lim V_1'(x) = 0, \quad \lim W_1(x) = \lim W_1'(x) = \infty,$$

and as x approaches $+\infty$

(13) $$\lim V_2(x) = \lim V_2'(x) = 0, \quad \lim W_2(x) = \lim W_2'(x) = \infty.$$

3. It is now necessary to consider how ϕ_1 and ϕ_2 vary with λ. If $u(x)$ is a solution expressed in the form (9) in which C and θ are independent of λ we may derive in the usual manner* the equation

(14) $$u'\partial u/\partial\lambda - u\partial u'/\partial\lambda = \int_0^x (\partial G/\partial\lambda)u^2dx,$$

since at the origin $\partial u/\partial\lambda = \partial u'/\partial\lambda = 0$ in view of (7). Let r be a root of $u(x)$, so that $-u_1(r)\sin\theta + u_2(r)\cos\theta = 0$. By differentiating this equation with respect to λ and making some simplifications by means of (8) and (14) we obtain

(15) $$\partial r/\partial\lambda = -\left[u_1^2(r) + u_2^2(r)\right]\int_0^x (\partial G/\partial\lambda)u^2dx.$$

This shows that as λ increases all roots of $u(x)$ move toward the origin. In the same manner if λ is constant and θ varies

(16) $$\partial r/\partial\theta = u_1^2(r) + u_2^2(r),$$

from which we see that as θ increases all roots of $u(x)$ move to the right. Finally if λ and θ vary in such a manner as to keep the root r fixed, we get from (15) and (16)

(17) $$\partial\theta/\partial\lambda = \int_0^r (\partial G/\partial\lambda)u^2dx.$$

We conclude that as λ increases θ also increases when r is positive, but decreases when r is negative. Now at a root of $u(x)$ we have $\theta = \phi(x) + k\pi$ $(k = 0, \pm 1, \pm 2, \cdots)$ so that

(18) $$\partial\phi/\partial\lambda = \int_0^r (\partial G/\partial\lambda)u^2dx.$$

From this we see that ϕ_2 increases as λ increases while ϕ_1 decreases as λ increases.

4. Now let λ increase from $-\infty$ to $+\infty$ and note the change in the quantity $\phi_2 - \phi_1$. Since there cannot be two roots of $u(x)u'(x)$ when λ is

* See Sturm, Journal de Mathématiques Pures et Appliquées, vol. 1 (1836), pp. 106–186, especially p. 113.

large and negative because $G(x, \lambda)$ is negative, it follows that, for such values of λ, $\phi_2 - \phi_1 < \pi$. But as λ increases there will be an infinite sequence of values of λ for which the increasing quantity $\phi_2 - \phi_1 = \pi, 2\pi, 3\pi, \cdots$, since $\phi_2 - \phi_1$ must increase without limit. We may denote these values of λ in order of increasing magnitude by $\lambda_0, \lambda_1, \lambda_2, \cdots$. For $\lambda = \lambda_n$, it is at once apparent that the solutions $V_1(x)$ and $V_2(x)$ are identical, except perhaps for sign, and the same is true of $W_1(x)$ and $W_2(x)$. We may therefore define a solution $U_n(x)$ corresponding to λ_n as follows:

$$U_n(x) = V_1(x) = \pm V_2(x), \text{ when } \lambda = \lambda_n. \tag{19}$$

We have therefore

$$\lim_{x=\pm\infty} U_n(x) = \lim_{x=\pm\infty} U_n'(x) = 0. \tag{20}$$

If we form equation (5) for the solution $U_n(x)$ and let x_2 approach $+\infty$ and x_1 approach $-\infty$, we see at once from (20) that the integrals

$$\int_{-\infty}^{+\infty} U_n'^2(x)dx \quad \text{and} \quad \int_{-\infty}^{+\infty} G(x,\lambda)U_n^2(x)dx$$

both converge. From the second of these we may conclude that

$$\int_{-\infty}^{+\infty} U_n^2(x)dx$$

also converges.

Since the number of roots of $U_n(x)$ is equal to $k-1$ when $\phi_2 - \phi_1 = k\pi$, we see that $U_n(x)$ has exactly n roots.

The foregoing conclusions may be summarized as follows:

THEOREM I. *There exists an infinite set of critical values of* λ, $\lambda_0, \lambda_1, \lambda_2, \cdots$, *with limit point at* $+\infty$ *only, corresponding to which equation* (1) *has solutions (unique except for a constant factor) satisfying the conditions*

$$\lim_{x=\pm\infty} U_n(x) = \lim_{x=\pm\infty} U_n'(x) = 0.$$

The solution $U_n(x)$ *vanishes* n *times in the interval* $-\infty < x < \infty$, *and the integrals*

$$\int_{-\infty}^{\infty} U_n^2(x)dx, \quad \int_{-\infty}^{\infty} U_n'^2(x)dx, \quad \text{and} \quad \int_{-\infty}^{\infty} G(x,\lambda)U_n^2(x)dx$$

all exist.

5. We return to the consideration of the conditions (4). The general solution of (1) may be written

$$u(x) = V_1(x)h + W_1(x)k,$$

so that the conditions (4) are equivalent to

$$
\begin{aligned}
&[\alpha_1 V_1(a) + \alpha_2 V_1'(a) + \alpha_3 V_1(b) + \alpha_4 V_1'(b)]h \\
&\qquad + [\alpha_1 W_1(a) + \alpha_2 W_1'(a) + \alpha_3 W_1(b) + \alpha_4 W_1'(b)]k = 0, \\
&[\beta_1 V_1(a) + \beta_2 V_1'(a) + \beta_3 V_1(b) + \beta_4 V_1'(b)]h \\
&\qquad + [\beta_1 W_1(a) + \beta_2 W_1'(a) + \beta_3 W_1(b) + \beta_4 W_1'(b)]k = 0.
\end{aligned} \tag{21}
$$

The determinant of these equations is

$$
\Delta = \begin{vmatrix}
[\alpha_1 V_1(a) + \alpha_2 V_1'(a) + \alpha_3 V_1(b) + \alpha_4 V_1'(b)] & \\
& [\alpha_1 W_1(a) + \alpha_2 W_1'(a) + \alpha_3 W_1(b) + \alpha_4 W_1'(b)] \\
[\beta_1 V_1(a) + \beta_2 V_1'(a) + \beta_3 V_1(b) + \beta_4 V_1'(b)] & \\
& [\beta_1 W_1(a) + \beta_2 W_1'(a) + \beta_3 W_1(b) + \beta_4 W_1'(b)]
\end{vmatrix}.
$$

We shall treat only the case for which A_{42} is not zero, as the modifications to be made when $A_{42}=0$ are sufficiently obvious. Let ϵ be a positive constant, as small as we please. We may choose b so large that $|V_1(b)/V_1'(b)|<\epsilon$, and then choose a so large (and negative) that $|V_1(a)|<\epsilon$, $|V_1'(a)|<\epsilon$, $W_1(a)/W_1'(a)<\epsilon$, $|W_1(b)/W_1'(a)|<\epsilon$, $|W_1'(b)/W_1'(a)|<\epsilon$, uniformly with respect to λ in an interval $\lambda_{n-1}+\epsilon \leqq \lambda \leqq \lambda_n-\epsilon$. Then the determinant may be written

$$\Delta = V_1'(b)W_1'(a)[A_{42}+\epsilon E],$$

in which E denotes a function that is bounded for λ in the given interval. Since $W_1'(a)$ does not vanish when a is large and negative and $V_1'(b)$ does not vanish for λ in the given interval when b is large, we see at once that Δ does not vanish in this interval. But in the next interval $\lambda_n+\epsilon \leqq \lambda \leqq \lambda_{n+1}-\epsilon$ the sign is changed and therefore Δ vanishes between $\lambda_n-\epsilon$ and $\lambda_n+\epsilon$. Therefore the roots of Δ approach λ_n $(n=0, 1, 2, \cdots)$. Moreover we can easily show that Δ does not vanish more than once in the interval $\lambda_n-\epsilon<\lambda<\lambda_n+\epsilon$, so that the characteristic numbers l_n of §1 are ultimately all simple and $\lim l_n=\lambda_n$.

From (12), (13), (19), and (20) it will be seen that (21) will be satisfied by taking $h\neq 0$, $k=0$ when a and b are infinite. Therefore the characteristic functions $u_n(x)$ of §1 approach the functions $U_n(x)$. We therefore have

THEOREM II. *As the ends of the interval recede to $-\infty$ and $+\infty$ the characteristic numbers all become simple and*

$$\lim l_n = \lambda_n \quad \lim u_n(x) = U_n(x) \qquad (n = 0,1,2, \cdots),$$

the limits being entirely independent of the boundary conditions (4).

UNIVERSITY OF OREGON,
EUGENE, ORE.

THE SECOND DERIVATIVE OF A POLYGENIC FUNCTION*

BY

EDWARD KASNER

In previous papers† I have studied the first derivatives of a polygenic function, especially from the geometric aspect. A polygenic function w (in opposition to an ordinary or monogenic function) of the complex variable $z=x+iy$, is any function of the form

$$w = \phi(x,y) + i\psi(x,y), \tag{1}$$

where the components ϕ and ψ are arbitrary functions (except for suitable continuity assumptions) of the real variables x, y, the Cauchy-Riemann conditions not being assumed.

The first derivative, which we denote by $\gamma=\alpha+i\beta$,

$$\gamma = dw/dz = \lim_{\Delta z \to 0} \Delta w/\Delta z$$

of course then depends not only on the point z but also on the direction of approach θ, that is, on the slope $m=y'=dy/dx$. The main theorem is that, plotted in the α, β plane, the locus of points γ, corresponding to a given point z, is a circle (*the derivative circle*‡), or more accurately a suitably parametrized circle, which I have termed a clock (*the derivative clock*).

Various expressions for γ are convenient, and are here given for reference:

$$\gamma = \frac{dw}{dz} = \frac{(\phi_x + i\psi_x) + m(\phi_y + i\psi_y)}{1 + im} = \frac{w_x + mw_y}{1 + im}; \tag{2'}$$

$$\gamma = \mathfrak{D}(w) + e^{-2i\theta}\,\mathfrak{P}(w), \tag{2''}$$

---

* Presented to the Society, April 6, 1928; received by the editors in February, 1928. I wish to express appreciation for the assistance of Miss Lulu Hofmann in writing this paper, especially in the discussion of the higher algebraic loci involved.

† *A new theory of polygenic functions*, Science, vol. 66 (1927), pp. 581–582; *General theory of polygenic or non-monogenic functions. The derivative congruence of circles*, Proceedings of the National Academy of Sciences, vol. 13 (1928), pp. 75–82; also L. Hofmann and E. Kasner, *Homographic circles or clocks* and E. Kasner, *Appendix on polygenic functions*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 495–503; E. Kasner, *Note on the derivative circular congruence of a polygenic function*, the same Bulletin, vol. 34, pp. 561–565, and two papers to appear in the Proceedings of the Bologna Congress.

‡ See a forthcoming paper by E. R. Hedrick, *On derivatives of non-analytic functions*, Proceedings of the National Academy of Sciences, where the derivative circle is termed the Kasner circle and new properties of it are studied.

where the linear operators

$$\mathfrak{D} = \tfrac{1}{2}(D_x - iD_y), \quad \mathfrak{P} = \tfrac{1}{2}(D_x + iD_y)$$

are introduced, the former defining mean differentiation, the latter the phase operator, so that $\mathfrak{D}(w)$ gives the center, and $\mathfrak{P}(w)$ the phase, of the derivative clock.

So far the real functions ϕ and ψ are assumed to be continuous and to have continuous partial derivatives. If we also assume ϕ and ψ to be analytic (that is, real power series in two variables x, y) we may consider the polygenic function w to be a power series (with complex coefficients) in the two variables (minimal coördinates)

$$z = x + iy \quad \text{and} \quad \bar{z} = x - iy;$$

and then it is easily seen that

$$\mathfrak{D}(w) = \partial w/\partial z = w_z, \quad \mathfrak{P}(w) = \partial w/\partial \bar{z} = w_{\bar{z}},$$

so that we have the simple formula*

$$\gamma = w_z + e^{-2i\theta} w_{\bar{z}}. \tag{2'''}$$

The present paper is devoted chiefly to the study of the second derivative d^2w/dz^2 of a polygenic function, and the appropriate geometry; higher derivatives and polygenic functions of more than one complex variable are considered only briefly in the final section. The formulas and the results are considerably more complicated. *The general second derivative depends on the curvature as well as the direction of approach.* Differentiating (2) with regard to z, that is, forming the ratio of the total differentials $d(dw/dz)$ and dz, we obtain

$$\frac{d^2w}{dz^2} = \frac{(1 + iy')(w_{xx} + 2w_{xy}y' + w_{yy}y'^2 + w_y y'') - (w_x + w_y y')iy''}{(1 + iy')^3}.$$

After simplifying this expression, and introducing for d^2w/dz^2 the complex notation $\sigma = \xi + i\eta$, we have

$$\sigma = \xi + i\eta = \frac{w_{xx} + 2w_{xy}y' + w_{yy}y'^2}{(1 + iy')^2} + \frac{w_y - iw_x}{(1 + iy')^3}y'' \equiv \Omega(x, y, y', y''). \tag{3}$$

The second derivative of a polygenic function w of z is thus a function of the point z at which it is formed and of the differential element of the second order (y', y'') along which this point is approached. Corresponding

* See G. Calugaréono, *Sur les fonctions polygènes d'une variable complexe*, Comptes Rendus, vol. 186 (1928), pp 930–932. N. Nicolesco, *Fonctions complexes dans le plan et l'espace*, Paris thesis, 1928. No geometry is given in these papers, and derivatives of higher order are considered only for rectilinear approach, while in the present paper the path of approach is allowed to be an arbitrary curve.

to the ∞^2 real elements of the second order existing at every point, d^2w/dz^2 assumes ∞^2 values for every value of z.* If we map these values of d^2w/dz^2 in the complex plane $\sigma=\xi+i\eta$, the mapping points will therefore fill out a region of this plane. It is not obvious, but will be proved as a special theorem later, that this region covers the entire plane, so that every point in the σ-plane corresponds to at least one real curvature element (y', y'') at the given point z.

The general character of the correspondence between the elements (y', y'') and the points (ξ, η) will be the same for all general points z. But of course as z varies, different points (ξ, η) will correspond to the same values of (y', y'').

We have seen that the second derivative of a polygenic function depends in general on y'' as well as on y'. If it is to be independent of y'', that is, if in (3) Ω is not to contain y'', the condition

$$w_y - iw_x = 0$$

has to be fulfilled. But this one complex equation splits up into the two Cauchy-Riemann equations, so that (the converse being evident) *the second derivative of a polygenic function is independent of y'' when and only when the function is monogenic*, which of course makes d^2w/dz^2 independent of y' also.†

Summary of results

In the following we shall first discuss the ∞^1 values assumed by d^2w/dz^2 at a definite point z as this point is successively approached along the various elements of the same constant curvature κ. The point mapping these values in the σ-plane describes an irrational curve of the eighth order as the slope of the considered element at z is varied continuously. For the special curvature $\kappa=0$, however, that is, for the approach of z along the straight line elements, this curve of the eighth order becomes a *limaçon* described twice.

After this, the corresponding problem for the elements at z tangent to a common fixed slope is discussed. The result is very simple. The locus of points of the σ-plane mapping the various values assumed by d^2w/dz^2 as z is approached along all the elements with the same slope, is a *straight line*; and this line is described (from a suitable initial point) at a rate proportional to the rate at which the curvature of the considered elements varies.

---

* We wish to stress the point that we consider only *real* elements of the second order just as in discussing the first derivative we considered only *real* elements of the first order.

† Similarly if we continue to higher derivatives of w, d^nw/dz^n will be a function of $x, y, y', y'', \cdots, y^{(n)}$, and will be independent of $y^{(n)}$ when and only when w is monogenic. So that when d^nw/dz^n is independent of $y^{(n)}$ it is also independent of $y', y'', \cdots, y^{(n-1)}$.

We then proceed to study the envelope of the ∞^1 lines corresponding in the just described manner to the various slopes at z. This envelope proves to be a *cardioid*.

Finally we investigate in two different ways the correspondence established by d^2w/dz^2 between the elements at z and the points of the σ-plane. It is found to be one-to-one in that region of the σ-plane formed by the points inside of the just defined cardioid, while to a point outside the cardioid correspond three curvature elements (y', y'').

The different representations of d^2w/dz^2

In the following we shall discuss exclusively the properties of d^2w/dz^2 at a fixed point (x, y) of the z-plane so that $w_x, w_y, w_{xx}, w_{xy}, w_{yy}$ are to be considered as certain complex constants and d^2w/dz^2 reduces to a function of y' and y'' only.

Since $d^2w/dz^2=\sigma$ is a function of the differential element of the second order along which the point z is approached, we can represent it in different ways by choosing different quantities to determine the element of approach. Formula (3) gives σ as a function of y' and y'',

$$\sigma = \Omega_1(y', y'').$$

By substituting into Ω_1,

$$y'' = (1 + y'^2)^{3/2} \cdot \kappa, \quad y' = \tan\theta = -i\frac{e^{i\theta} - e^{-i\theta}}{e^{i\theta} + e^{-i\theta}},$$

we obtain a representation of σ as a function of the curvature κ and the direction angle θ of the element of approach,

$$\sigma = \tfrac{1}{4}[w_{xx} - 2iw_{xy} - w_{yy} + 2(w_{xx} + w_{yy})e^{-2i\theta} + (w_{xx} + 2iw_{xy} - w_{yy})e^{-4i\theta}] - i(w_x + iw_y)e^{-3i\theta} \cdot \kappa.$$

The coefficients in this expression simplify, if instead of x and y we use $z=x+iy$ and $\bar{z}=x-iy$ as independent variables. Then

$$w_x - iw_y = 2w_z, \quad w_x + iw_y = 2w_{\bar{z}}, \quad w_{xx} - 2iw_{xy} - w_{yy} = 4w_{zz}$$

and so forth, so that finally

$$\sigma = w_{zz} + 2w_{z\bar{z}}e^{-2i\theta} + w_{\bar{z}\bar{z}}e^{-4i\theta} - 2iw_{\bar{z}}e^{-3i\theta} \cdot \kappa \equiv \Omega_2(\theta, \kappa).$$

A differential element of the second order at the point (x, y) is also determined when the center of its circle of curvature is given. If X and Y denote the coördinates of this center relative to the point (x, y), the well known transformation formulas are

$$y' = -\frac{X}{Y}, \quad y'' = \frac{X^2+Y^2}{Y^3}.$$

The substitution of these expressions for y' and y'' into Ω_1 furnishes

$$\sigma = \frac{w_{xx}Y^2 - 2w_{xy}XY + w_{yy}X^2 + (w_y - iw_x)(Y + iX)}{(Y - iX)^2} \equiv \Omega_3(X, Y).$$

By using as before z and $\bar{z}$ as independent variables and correspondingly $Z = X + iY$ and $\overline{Z} = X - iY$, we obtain

$$\sigma = \frac{w_{zz}Z^2 - 2w_{z\bar{z}}Z\overline{Z} + w_{\bar{z}\bar{z}}\overline{Z}^2 + 2w_z\overline{Z}}{Z^2} \equiv \Omega_4(Z, \overline{Z}).$$

The main importance of the two last representations of σ lies in the fact that they convert the element-point correspondence between the elements at the point z and the points of the σ-plane expressed by Ω_1 and Ω_2 into a point correspondence between the Z-plane and the σ-plane.

For the sake of easier reference, we put the four different representations of σ that we have developed together:

$$\text{(A)} \quad \begin{cases} \sigma = \dfrac{w_{xx} + 2w_{xy}y' + w_{yy}y'^2}{(1 + iy')^2} + \dfrac{(w_y - iw_x)y''}{(1 + iy')^3} \equiv \Omega_1(y', y''), \\ \sigma = w_{zz} + 2w_{z\bar{z}}e^{-2i\theta} + w_{\bar{z}\bar{z}}e^{-4i\theta} - 2iw_z e^{-3i\theta} \cdot \kappa \equiv \Omega_2(\theta, \kappa), \\ \sigma = \dfrac{w_{xx}Y^2 - 2w_{xy}XY + w_{yy}X^2 + (w_y - iw_x)(Y + iX)}{(Y - iX)^2} \equiv \Omega_3(X, Y), \\ \sigma = \dfrac{w_{zz}Z^2 - 2w_{z\bar{z}}Z\overline{Z} + w_{\bar{z}\bar{z}}\overline{Z}^2 + 2w_z\overline{Z}}{Z^2} \equiv \Omega_4(Z, \overline{Z}). \end{cases}$$

THE RECTILINEAR SECOND DERIVATIVE

As the fixed point z is approached along the different elements belonging to a certain definite slope, d^2w/dz^2 assumes ∞^1 values. Among these we distinguish the value corresponding to the approach of z along the element of curvature $\kappa = 0$, that is, along the straight line element of that particular slope, and term it the *rectilinear second derivative* of w belonging to that slope. We abbreviate it by σ_0, so that

$$\sigma_0 = \frac{w_{xx} + 2w_{xy}y' + w_{yy}y'^2}{(1 + iy')^2} \equiv \Omega_1(y', 0)$$
$$= w_{zz} + 2w_{z\bar{z}}e^{-2i\theta} + w_{\bar{z}\bar{z}}e^{-4i\theta} \equiv \Omega_2(\theta, 0).$$

The rectilinear second derivative is a function of the slope y' or of the direction angle θ. As y' varies from 0 to ∞, that is, as θ varies from 0 to π, it assumes ∞^1 values, and the point σ_0 mapping these values in the complex

plane σ describes a curve. This curve, which we call L, is most easily discussed by means of the second representation of σ_0:

$$\sigma_0 = w_{zz} + w_{\bar{z}\bar{z}}e^{-4i\theta} + 2w_{z\bar{z}}e^{-2i\theta},$$

in which now θ is a variable going from 0 to π.

A point σ_0 of L is determined as the sum of the three vectors w_{zz}, $w_{\bar{z}\bar{z}}e^{-4i\theta}$ and $2w_{z\bar{z}}e^{-2i\theta}$. As θ varies, w_{zz} remains fixed and the end points of $w_{\bar{z}\bar{z}}e^{-4i\theta}$ and $2w_{z\bar{z}}e^{-2i\theta}$ separately describe circles of radii $|w_{\bar{z}\bar{z}}|$ and $2|w_{z\bar{z}}|$ with the angular velocities -4θ and -2θ. Since the vector $2w_{z\bar{z}}e^{-2i\theta}$ is applied at the end point of the vector $w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta}$ and rotates with one half the angular velocity of $w_{\bar{z}\bar{z}}e^{-4i\theta}$, the lines S_1 and S_2 carrying any two vectors $2w_{z\bar{z}}e^{-2i\theta_1}$ and $2w_{z\bar{z}}e^{-2i\theta_2}$ applied as described at the points $\sigma_1=w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta_1}$ and $\sigma_2=w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta_2}$ must intersect at a fixed point P of the circle $\sigma=w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta}$; for these lines form the angle $-2(\theta_2-\theta_1)$ and this is equal to one half of the central angle $-4(\theta_2-\theta_1)$ over the points σ_1 and σ_2 of the circle which lie on the lines S_1 and S_2 respectively according to the construction. It follows from this discussion that the curve L is the locus of the points lying on the lines through the point P of the circle $\sigma=w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta}$ at the distance $2|w_{z\bar{z}}|$ in either direction from the second point of intersection of these lines with the circle (in *either* direction since θ goes from 0 to π and to the angles θ and $\theta+\pi/2$ correspond the points $\sigma_0=w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta}\pm 2w_{z\bar{z}}e^{-2i\theta}$). This however is the ordinary definition of a *limaçon* with the basal circle $\sigma=w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta}$, the pole P, and the determining length $2|w_{z\bar{z}}|$.*

Our discussion not only characterises the curve L as a whole but also indicates its parametrization with regard to θ. Let us distinguish on the limaçon the point Q: $\sigma_0=w_{zz}+w_{\bar{z}\bar{z}}+2w_{z\bar{z}}$ corresponding to $\theta=0$ as an initial point. Then given the limaçon, the pole P and the point Q, the point corresponding to an arbitrary value of θ is found by turning the line PQ about P through the angle -2θ, as that point of the limaçon into which this continuous motion carries the point of intersection of the rotating line with the limaçon.

The values of the rectilinear second derivative of a polygenic function w belonging to the different slopes at the point z are represented in the complex plane σ of d^2w/dz^2 by the points of a limaçon L. When the direction angle of the slope at z varies, the corresponding point on the limaçon moves with twice the angular velocity and in the opposite direction, the angular velocity being measured from the pole of the limaçon as center.

---

* See Gino Loria, *Spezielle algebraische und transzendente ebene Kurven*, Leipzig, Teubner, 1902, p. 136.

SPECIAL TYPES OF POLYGENIC FUNCTION

The limaçon L specializes in certain noteworthy simple cases.

(a) When $w_{zz}=0$, so that

$$w = f_1(z) + f_2(\bar{z}),$$

that is, what I have termed a *harmonic* polygenic function,* then L becomes a clock of rate -4.† As θ varies from 0 to π, the corresponding point σ_0 describes this clock twice, assuming equal values for two angles θ differing by $\pi/2$.

(b) When $w_{\bar{z}\bar{z}}=0$,

$$w = f_1(z) + \bar{z}f_2(z),$$

and L becomes a clock of rate -2.

(c) When $w_{z\bar{z}}=0$,

$$w = f_1(\bar{z}) + zf_2(\bar{z}).$$

Then L remains a limaçon but is so far specialised in position as to have the center of its basal circle at the origin.

(d) When $w_{z\bar{z}}=0$ and $w_{\bar{z}\bar{z}}=0$,

$$w = f(z) + A\bar{z},$$

and L reduces to a point.

(e) When $w_{\bar{z}\bar{z}}=0$ and $w_{zz}=0$,

$$w = A + Bz + C\bar{z} + Dz\bar{z},$$

and L becomes a clock of rate -2 with its center at the origin.

(f) When $w_{zz}=0$ and $w_{z\bar{z}}=0$,

$$w = f(\bar{z}) + Az$$

and L becomes a clock of rate -4 with center at the origin.

(g) When $w_{zz}=0$, $w_{z\bar{z}}=0$ and $w_{\bar{z}\bar{z}}=0$, $w = A + Bz + D\bar{z}$, and L reduces to the point at the origin.

(h) When $|w_{z\bar{z}}| = |w_{\bar{z}\bar{z}}|$, that is when the determining length $2|w_{z\bar{z}}|$ of the limaçon is equal to the diameter $2|w_{\bar{z}\bar{z}}|$ of its basal circle, the limaçon L specialises into a cardioid.‡

The converses of these eight theorems are also true, so that the geometric properties stated are completely characteristic of these types of polygenic function.

---

* A polygenic function $w=F(z)=\phi(x, y)+i\psi(x, y)$ is called harmonic when ϕ and ψ separately fulfill the Laplace equation. See Proceedings of the National Academy of Sciences, loc. cit., p. 81. The mean derivative of such a polygenic function is a monogenic function.

† The circle $\sigma_0=w_{zz}+w_{\bar{z}\bar{z}}e^{-4i\theta}$ is parametrized with regard to θ in such a manner that as θ varies the point σ_0 on the circle moves with four times the angular velocity and in the opposite sense.

‡ A cardioid is defined as a limaçon for which the determining length is equal to the diameter of the basal circle. See Loria, loc. cit., p. 142.

THE CURVILINEAR SECOND DERIVATIVE CORRESPONDING TO THE ELEMENTS OF A FIXED CURVATURE

After having studied the curve described by σ_0,

$$\sigma_0 = \Omega_2(\theta, 0),$$

the points of which map the values of the rectilinear second derivative corresponding to the different slopes at the point z, let us now study the curve described by the point

$$\sigma_\kappa = w_{zz} + 2w_{zz}e^{-2i\theta} + w_{\bar{z}\bar{z}}e^{-4i\theta} - 2iw_{\bar{z}}e^{-3i\theta}\cdot\kappa \tag{4}$$

as θ varies and κ remains constant. The point σ_κ maps the values of d^2w/dz^2 corresponding to the different elements of the fixed curvature κ at z. If we call the value assumed by d^2w/dz^2 as the point z is approached along an element of a certain slope and a certain curvature $\kappa \neq 0$ the *curvilinear second derivative of* w corresponding to that slope and curvature, then σ_κ maps the values of the curvilinear second derivative corresponding to the various elements of the curvature κ.

Obviously

$$\sigma_\kappa = \sigma_0 - 2iw_{\bar{z}}\kappa e^{-3i\theta}.$$

The term added to σ_0 represents a point describing the circle of radius $2\kappa|w_{\bar{z}}|$ about the origin at the rate -3 as θ varies. The points σ_κ can thus be determined as the vectorial sum of a point describing the limaçon L at the rate -2 in the sense explained above, and a point describing a certain circle about the origin at the rate -3.

The resulting curve is not rational. This is seen by substituting into $\sigma = \Omega_1(y', y'')$

$$y'' = (1 + y'^2)^{3/2}\kappa.$$

Then in

$$\sigma_\kappa = \xi_\kappa + i\eta_\kappa = \Omega_1(y', [1 + y'^2]^{3/2}\cdot\kappa),$$

ξ_κ and η_κ are algebraic but not rational functions of y', since the irrational expression $(1+y'^2)^{1/2}$ enters essentially and cannot be eliminated.

To determine the order of the curve described by σ_κ it is necessary to note that if this is to be a complete closed curve, we must let θ vary from 0 to 2π on account of the odd exponent -3 of the last term in (4). The reason is obvious since as θ goes from 0 to π only, the element $(0, \kappa)$ at z is carried by a continuous rotation into the element (π, κ) and these two elements are not the same but lie on opposite sides of their tangent. Generally this variation of θ would not furnish all elements of the curvature κ, but for every

slope only that one element into which the above described rotation carries $(0, \kappa)$.†

Equation (4) can be written in the form

$$\sigma_\kappa - w_{zz} = w_{zz}e^{-2i\theta}(e^{-i\theta} + A)(e^{-i\theta} + B),$$

where A and B are determined by w_{zz}, $2w_{z\bar{z}}$ and $-2iw_z\kappa$. This expression has an algebraic limiting case

$$\sigma^* = e^{-4i\theta}.$$

Here the point σ^* describes the unit circle about the origin four times as θ goes from 0 to 2π, so that the curve as a whole must be counted as a curve of the eighth order. *The curve* (4) *therefore will be at least of order eight.* That it is at most of order eight is shown as follows. Replace, in (4), $e^{-i\theta}$ by $\zeta = \zeta_1 + i\zeta_2$, and assume for a moment that ζ_1 and ζ_2 are independent variables instead of making them fulfill the relation $\zeta_1^2 + \zeta_2^2 = 1$ which makes $e^{-i\theta} = \zeta$. Then $\sigma_\kappa = \xi_\kappa + i\eta_\kappa$ assumes ∞^2 values as ζ_1 and ζ_2 vary and we have a point correspondence between the ξ-plane and the σ_κ-plane which is superposed over the σ-plane. Splitting up into real and imaginary parts, we see that ξ_κ and σ_κ are integral functions of the fourth degree of ζ_1 and ζ_2, so that to every line of the σ_κ-plane corresponds a certain curve of the fourth degree in the ζ-plane. Now the curve described by σ_κ according to (4), is equal to that curve of the σ_κ-plane into which the transformation just described carries the circle $\zeta_1^2 + \zeta_2^2 = 1$ of the ζ-plane. This circle is cut by one of the curves of the fourth order corresponding to the lines of the σ_κ-plane in at most eight points that vary with the chosen curve. Therefore, in virtue of the correspondence, the curve described by σ_κ is intersected by a line of its plane in at most eight points. Thus in view of the fact that the order of this curve was above shown to be at least eight, it is now proved to be exactly eight.

The values of the curvilinear second derivative of a polygenic function w corresponding to the elements of a fixed curvature κ at the point z are mapped in the complex plane σ of d^2w/dz^2 by the points of an irrational curve of the eighth order.‡ The points of this curve are determined as the vectorial sum of a point

† The variation of θ from 0 to π only is sufficient to furnish all elements of the curvature κ only in the case $\kappa = 0$.

‡ The seeming contradiction that for a general value of the curvature κ, σ_κ describes a curve of the eighth order while for the special value $\kappa = 0$, σ_0 describes a limaçon, i.e. a curve of the fourth order, is solved by the remark that the limaçon was obtained by varying θ from 0 to π only. If also in the case $\kappa = 0$ we let θ go from 0 to 2π, σ_0 will describe the limaçon twice so that the entire curve will here also be a curve of the eighth order.

of the limaçon L and a point of a certain circle about the origin, the point on the limaçon moving at the rate −2 in the sense explained above as the slope at z varies, and the point on the circle at the rate −3.

THE CURVILINEAR SECOND DERIVATIVE CORRESPONDING TO THE ELEMENTS WITH A FIXED COMMON TANGENT

So far we have examined the points mapping the values assumed by d^2w/dz^2 for the approach of the point z along the elements of a certain fixed curvature. We shall now investigate the case in which z is approached successively along the elements of a certain fixed slope, that is, along the elements with a common tangent. That means that we must study the curve described by the point

$$\sigma_\theta = (w_{zz} + 2w_{z\bar{z}}e^{-2i\theta} + w_{\bar{z}\bar{z}}i^{-4i\theta}) - 2iw_{\bar{z}}e^{-3i\theta}\cdot\kappa, \tag{5}$$

as κ varies from $-\infty$ to $+\infty$ and θ remains fixed; or, using the representation of σ by means of $\Omega_1(y', y'')$ and abbreviating y' by m, the curve described by the point

$$\sigma_m = \frac{w_{zz} + 2w_{zy}m + w_{yy}m^2}{(1+im)^2} + \frac{w_y - iw_z}{(1+im)^3}\cdot y'', \tag{6}$$

as y'' varies from $-\infty$ to $+\infty$ and $m = \tan\theta$ remains fixed.

Since the expression on the right of (5) is an integral linear function of κ, it follows that *as κ varies from $-\infty$ to $+\infty$, σ_θ describes a straight line and describes this line at a rate proportional to the rate at which κ varies.* We abbreviate this line by S_θ or $S_m(m = \tan\theta)$.†

As (5) shows, S_θ is parallel to the line carrying the vector $-2iw_{\bar{z}}e^{-3i\theta}$. If therefore we construct two lines S_{θ_1} and S_{θ_2} which map the values of d^2w/dz^2 corresponding once to the elements with the common tangent of slope θ_1, once to the elements with the common tangent of slope θ_2, S_{θ_1} and S_{θ_2} form the angle $-3(\theta_2-\theta_1)$, that is -3 times the angle $(\theta_2-\theta_1)$ formed by the determining slopes at the point z. From the fact that when θ varies the direction angle of S_θ varies by the triple amount and in the opposite sense, it follows that there are four real slopes θ at z which are parallel to the corresponding lines S_θ. The proof of this mainly comes down to solving the equation

$$\tan(\alpha + \theta) = -\tan(\beta + 3\theta),$$

where α and β are constant angles, for $\tan\theta$. Obviously this equation is of

† As κ varies it naturally also assumes the value $\kappa = y'' = 0$. Thus the line L_m goes through that point of the limaçon L which maps the value of the rectilinear derivative belonging to the slope m.

the fourth degree in $\tan\theta$; a short geometric consideration shows that the four roots are real.

The values of the curvilinear second derivative of a polygenic function w corresponding to the elements with a fixed common tangent of slope θ at the point z, are mapped in the complex plane σ of d^2w/dz^2 by the points of a straight line S_θ or S_m. As the curvature of the elements at z varies, the mapping point describes the line S_θ at a rate proportional to the rate of variation of the curvature itself. As the direction angle θ of the slope at z varies, the direction angle of the corresponding line S_θ varies by the triple quantity and in the opposite sense. There are four real slopes at z which are parallel to the corresponding lines S_θ in the σ-plane.

The envelope of the lines L_θ

After discussing the individual lines L_θ we proceed to consider them as a totality and to study the curve which they envelop as θ goes continuously from 0 to π. This envelope will be called E. To determine the character of E it is necessary to obtain the real equation of the lines L_θ. For this purpose we use the representation of σ by means of (6) and abbreviate it by

$$\sigma_m = \sigma_m\xi_m + i\eta_m = a(m) + b(m)y''. \tag{6'}$$

Let us split a and b into their real and imaginary parts $a=a_1+ia_2$ and $b=b_1+ib_2$. According to (6'), S_m has the slope b_2/b_1 and goes through the point $(a_1, a_2)(a_1+ia_2=\Omega_2(\theta, 0), \theta=\arctan m)$; it therefore has the equation

$$(\eta_m - a_2) / (\xi_m - a_1) = b_2 / b_1.$$

We write this equation in the form

$$L_m \equiv p\xi + q\eta + r = 0\,;$$

a simple calculation furnishes the following values for p, q and r:†

$$\begin{aligned} p &= Am^3 - 3Bm^2 - 3Am + B, \\ q &= Bm^3 + 3Am^2 - 3Bm - A, \\ r &= m[Bg(m) + Af(m)] + Ag(m) - Bf(m), \end{aligned} \tag{7}$$

where

$$A = \phi_y + \psi_x,\ B = -\phi_x + \psi_y,$$
$$f(m) = \phi_{xx} + 2m\phi_{xy} + m^2\phi_{yy},\ g(m) = \psi_{xx} + 2m\psi_{xy} + m^2\psi_{yy}.$$

By (7) the three homogeneous line coördinates of S_m are given as functions of the parameter m. As A and B are constants (only depending on the point z) and $f(m)$ and $g(m)$ are polynomials of the second degree in m, the three homogeneous line coördinates p, q, r of L_m are integral rational functions of the

† For the sake of greater convenience we here and in the sequel omit the factor of proportionality by which the homogeneous coördinates p, q, and r should be multiplied in the parametric representation (7).

third degree in the parameter m. The curve E was defined as the envelope of the lines L_m. Therefore (7), as the parametric representation of these lines, is the parametric representation of E as locus of its tangents. It follows from the rationality and the degree three of p, q, and r with regard to m that the envelope E *is a rational curve of class three* (the genus of a curve is the same whether the curve be considered as locus of its tangents or of its points).

To determine the order of E we introduce the homogeneous point coordinates ξ_m^*, η_m^* and t for the points of E,

$$\xi_m^* / t = \xi_m, \ \eta_m^* / t = \eta_m,$$

and derive the parametric representation of E as locus of its points from (7) in the familiar manner. We have

$$(8) \qquad \xi^* = qr' - q'r, \ \eta^* = rp' - r'p, \ t = pq' - p'q,$$

where the prime denotes differentiation with regard to the parameter m.

A rational curve of the third class is either of the fourth or of the third order.† We easily find

$$(9) \qquad t = -3(A^2 + B^2)(1 + m^2)^2.$$

Since t is actually of the fourth degree, the curve E can be of a lower order only if ξ^*, η^* and t have some polynomial in m as a common factor. According to (9) this factor could only be $m-i$ or $m+i$. A short calculation however shows that ξ^* and η^* do not vanish for $m=i$ or $m=-i$. Thus it is proved that E *is of the fourth order*.

The singularities of a rational curve of the fourth order and the third class consist in an ordinary double tangent and three ordinary cusps of which one is real and two conjugate imaginary.† Without going into details about the double tangent or the real cusp, we shall merely determine the position of the imaginary cusps of E. The discussion which we give in detail later on shows these cusps to lie at *the cyclic points* of the σ-plane.

Under the assumption that this is proved, E *is now characterised as a rational curve of order four and class three with its imaginary cusps at the cyclic points*. This characterisation is sufficient to show that E is a well known curve of very special type, namely a *cardioid*. The proof is as follows.

There are ∞^3 rational curves of order four and class three.‡ The require-

† This follows immediately from the consideration of the well known dual curve, the rational curve of the third order. A rational curve of the third order has either an ordinary double point or a cusp. In the first case it is of class four, in the second of class three. Furthermore a rational curve of the third order and fourth class has one real and two imaginary inflexion tangents, which together with its double point make up all its singularities.

‡ This again follows from the fact that there are ∞^8 dual curves of order three and class four. For proof see Severi-Loeffler, *Vorlesungen ueber Algebraische Geometrie*, Leipzig, 1921, p. 161–162.

ment that these curves shall have their imaginary cusps at the cyclic points (or at any two definite points) represents a fourfold condition. Thus there are ∞^4 rational curves of order four and class three with their imaginary cusps at the cyclic points. We abbreviate this family of curves by (E). On the other hand the properties characterising the curves (E) are, as is well known, properties of the cardioids, so that the family of cardioids must be contained among the curves (E). The cardioids themselves, however, form a four-parameter family of curves, a general cardioid containing one parameter of size and three of position. Then since the dimension of the family (E) is equal to the dimension of the family of cardioids and at the same time contains this family, the curves (E) must be identical with the cardioids.

Hereby it is proved that *the envelope E as a curve of* (E) *is a cardioid.*

We now give the discussion referring to the imaginary cusps of E which was postponed above.

Let us form the expression

$$i\xi_m^* + \eta_m^* + \lambda t, \tag{10}$$

where λ is an arbitrary constant. Substituting for ξ_m^*, η_m^* and t their parametric expressions in m according to (8) and (7), where, however, t need not be written out in full, we find for (10)

$$(m-i)^2[(A-iB)\{-(m-i)r'+3r\}-3\lambda(A^2+B^2)(m+i)^2]. \tag{10'}$$

The equation obtained by putting (10) equal to zero represents a condition on the coördinates ξ_m^*, η_m^*, t; it will be fulfilled by the coördinates of only those points of E which lie on the line

$$i\xi + \eta + \lambda = 0,$$

that is, a certain line through the cyclic point $I_1(\xi/\eta = i)$ of the σ-plane. The values of the parameter m determining these points of E are of course obtained as the roots of (10′) put equal to zero. But this equation obviously always has the double root $m=i$, no matter what value λ is given. That means that all the lines through I_1 cut E twice in the same fixed point, namely the point corresponding to the parameter value $m=i$. Since, however, the only point common to all the lines through I_1 is I_1 itself, it follows that E must have this point for a double point.

The discussion for $I_2(\xi/\eta = -i)$ is analogous. Since we know from the preceding arguments that every double point of E is a cusp, it is thus proved that E has cusps at the cyclic points of the σ-plane.

Combining the results of this paragraph with those relating to the ndividual lines L_m previously obtained, we have the following theorem:

As the slope m at the point z varies, the corresponding line L_m in the σ-plane moves so that it envelops a cardioid E in general position. Given the cardioid and, as an initial tangent, the line L_0 corresponding to the slope $m=0$ at z, the tangent corresponding to an arbitrary value $m=\tan\theta$ of the slope m is determined by letting the initial tangent glide along the cardioid in the negative sense till its direction angle has grown by the quantity -3θ. There are four tangents to the cardioid parallel to the determining slopes at z.

THE CORRESPONDENCE BETWEEN THE CURVATURE ELEMENTS AT z AND THE POINTS OF THE σ-PLANE

By means of the formula

$$\sigma = \Omega_1(y', y'')$$

a certain correspondence is set up between the ∞^2 real elements (y', y'') at the point z and the points of the σ-plane. Since Ω_1 is a rational function of y' and y'', it follows that to every element (y', y'') corresponds one and only one point σ. The converse question, how many elements (y', y'') correspond to one point σ, is answered by the following consideration.

We have in these last paragraphs grouped the ∞^2 real elements of the second order at z into ∞^1 groups each of which contained all the elements tangent to the same slope. The corresponding grouping of the points σ was that according to the real tangents of the cardioid E.† Thus a point σ represents as many elements (y', y'') as there are real tangents to E going through it. And since a cardioid is a curve of the third class, this number is equal to one for a point in the inside of E and equal to three for every other point.‡

By means of the second derivative d^2w/dz^2 of a polygenic function w of z, a certain rational correspondence is set up between the real elements of the second order at a definite point z and the points σ mapping the values assumed by d^2w/dz^2 as z is approached along these elements. This correspondence is one-to-one in that region of the σ-plane which is formed by the points inside the cardioid E and three-to-one in the rest of the σ-plane. Thus every point σ at the inside of E maps one value and every other point σ maps three values of d^2w/dz^2.

Since every point of the σ plane maps at least one value of d^2w/dz^2, the statement made at the very beginning of this paper is verified, that *as d^2w/dz^2 is calculated at a point z for all possible real elements (y', y'') of approach, the mapping point σ covers the entire σ-plane.*

† While discussing the lines L_θ we did not stress the fact that they are real, as being obvious since we consider only real values of θ and κ.

‡ These tangents are distinct for a general point σ at the outside of E. For those points lying on the double tangent of E or on E itself, two of them coincide; for the cusp of E and for the points of contact of the double tangent, all three coincide.

THE POINT CORRESPONDENCE BETWEEN THE Z- AND THE σ-PLANE

The results of the preceding paragraph can also be obtained without taking the cardioid E into consideration. So far we have made use only of the first two representations of σ in system (A), p. 807. We now turn to the remaining ones:

$$(11)\quad \sigma = \frac{w_{xx}Y^2 - 2w_{xy}XY + w_{yy}X^2 + (w_y - iw_x)(Y + iX)}{(Y - iX)^2} \equiv \Omega_3(X, Y),$$

$$(11')\quad \sigma = \frac{w_{zz}Z^2 - 2w_{z\bar{z}}Z\bar{Z} + w_{\bar{z}\bar{z}}\bar{Z}^2 + 2w_z\bar{Z}}{Z^2} \equiv \Omega_4(Z, \bar{Z}).$$

These formulas express a point correspondence between the (X, Y)- or $X + iY = Z$-plane and the $\xi + i\eta = \sigma$-plane, where X and Y are the center coördinates relative to the point z of the circle of curvature of the element (y', y'') at z.

Since in (11) σ is a rational function of X and Y there will be one point σ corresponding to every real point (X, Y). We ask the converse question, how many real points (X, Y) correspond to the same point σ. This is identical with the question of how many values of d^2w/dz^2 are represented by the same point σ, since every point (X, Y) determines one element (y', y'') which again determines one value of d^2w/dz^2.

To solve our problem, we write (11) in the form

$$(12)\quad (\xi + i\eta)(Y - iX)^2 = w_{xx}Y^2 - 2w_{xy}XY + w_{yy}X^2 + (w_y - iw_x)(Y + iX),$$

where ξ and η are now assumed to have some definite values and X and Y vary. Equation (12) represents a complex conic in the (X, Y)-plane. Therefore there will be as many real points (X, Y) corresponding to the fixed point (ξ, η), i.e., points (X, Y) whose coördinates fulfill equation (12), as there are real points on the conic represented by (12). From this number, however, we must naturally exclude such fixed real points as are common to all conics (12) independently of ξ and η. Now the number of real points on (12) is equal to the number of real intersection points of the two real conics obtained by equating the real and imaginary parts of (12) separately to zero. Since the total number of intersection points of these two conics is four, and since they have the fixed intersection $X = Y = 0$ independently of (ξ, η), the number of their variable real intersection points is equal either to three or to one.

The discussion as to which points (ξ, η) correspond to a complex conic (12) with three real variable points and which to one with one real variable point would of course lead to the cardioid E in σ. We will not carry it through but merely state the complete theorem as follows:

By means of the second derivative d^2w/dz^2 of a polygenic function w of z a certain rational correspondence is established between the points Z indicating the centers of the curvature circles of the elements (y', y'') at a fixed point z and the points σ mapping the values of d^2w/dz^2 as z is approached along these elements. This correspondence is one-to-one in that region of the σ-plane formed by the inside points of the cardioid E and three-to-one in the remainder of the σ-plane, so that to a point σ in the first region corresponds one and to any other point σ correspond three points Z.

Derivatives of higher order

The nth derivative of a polygenic function is a complex quantity depending on the given point x, y and on the derivatives $y', y'', \cdots, y^{(n)}$ of the curve of approach. If the highest derivative $y^{(n)}$ is to be absent it turns out that all the other derivatives are absent and the function is necessarily monogenic.

The general form of expression for *the third derivative* of any polygenic function is found to be

$$d^3w/dz^3 = \beta_1 + \beta_2 y'' + \beta_3 y''^2 + \beta_4 y''',$$

where the coefficients involve x, y, y'. Many geometric theorems may be deduced from this. We mention only the simplest, namely that the locus corresponding to fixed values y', y'', and variable y''' is a straight line.

The fourth derivative is

$$d^4w/dz^4 = (B_1 + B_2 y'' + B_3 y''^2 + B_4 y''^3) + (B_5 + B_6 y'') y''' + B_3 y'''',$$

which is seen to be linear in the two highest derivatives y''', y''''. Analogous results are found for all higher derivatives but we shall not take up the detailed theory here.

Polygenic functions of two or more independent complex variables will be studied in another paper. If w is a polygenic function of $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, the two partial derivatives of first order $P = \partial w/\partial z_1$ and $Q = \partial w/\partial z_2$ give rise to a pair of circles or clocks (for any given pair of points z_1 and z_2). But on account of the *essentially new feature* $\partial^2 w/\partial z_1 \partial z_2 \neq \partial^2 w/\partial z_2 \partial z_1$ (equality holding in fact only when w is an analytic function), there are *four* distinct partial derivatives of second order. $R = \partial^2 w/\partial z_1^2$, $S = \partial^2 w/\partial z_1 \partial z_2$, $S^* = \partial^2 w/\partial z_2 \partial_1$, $T = \partial^2 w/\partial z_2^2$, and the corresponding geometry is of fascinatingly complicated structure.

Columbia University,
New York, N. Y.

DIOPHANTINE EQUATIONS IN DIVISION ALGEBRAS*

BY

RALPH G. ARCHIBALD

1. **Introduction.** A type of division algebras of order n^2, defined over a field F,—called algebras of type D—has been given by Professor L. E. Dickson.† A further division algebra of order sixteen has lately been given by Professor F. Cecioni.‡ A great advance in the problem of the determination of all division algebras has been made in Dickson's recent paper on *New division algebras*.§ In that paper it is shown how to construct algebras Γ of order p^2q^2 over a field F, corresponding to the Galois group, of order $n=pq$, of an equation $f(x)=0$, of degree n, irreducible in the field F. General conditions that algebras Γ be associative are derived; and, in particular, associativity conditions are deduced for algebras Γ whose group for the field F is abelian and has two generators. In §12 of his paper Dickson gives necessary and sufficient conditions, for the case $p=q=2$, that an algebra Γ, whose group is abelian for the field F, be an associative division algebra.

It is the object of the present paper to obtain necessary and sufficient conditions for the solvability, in integers, of the diophantine equations which arise in an attempt to satisfy the associativity conditions for an algebra Γ based on the quartic equation $x^4+px^2+n^2=0$ irreducible in the field of rational numbers. In §2 it is noted that, in view of a known theorem on Tschirnhausen transformations, quite a general situation is being considered when the quartic equation is assumed to be of the above form. In §3 a diophantine equation in six unknowns is obtained which, by simple change of variables (in §4), simplifies to

$$U^2 - a_1R_0^2 = (A_0^2 - a_1b_1B_0^2)(F_1^2 - b_1C_0^2),$$

where each of $a_1=2n-p$, $b_1=-2n-p$, a_1b_1 is an integer different from a perfect square. In §5, by the use of Dickson's result on the integral solutions of $x^2-my^2=zw$, the nature of the problem is modified; in §7 necessary and sufficient conditions for the solvability, in integers, of the diophantine equation are obtained and simplified; and, finally, in §9 numerical examples

* Presented to the Society, October 29, 1927; received by the editors in December, 1927.

† *Algebras and their Arithmetics*, §47, p. 65.

‡ Rendiconti del Circolo Matematico di Palermo. vol. 47 (1923), pp. 209–254.

§ These Transactions, vol. 28 (1926), pp. 207–234.

are given which, incidentally, show that the conditions obtained are not inconsistent.

2. **Quartic equation irreducible in the field of rational numbers.** It is known* that, by means of a rational quadratic Tschirnhausen transformation, a quartic equation with rational coefficients, irreducible in the field of rational numbers, whose Galois group for the field of rational numbers is the group

$$G_4 = \{I, (12)(34), (13)(24), (14)(23)\},$$

can be brought to the form

$$x_1^4 + p_1 x_1^2 + n_1^2 = 0,$$

where p_1 and n_1 are rational numbers.

Write $n_1 = n_2/n_3$, $p_1 = p_2/p_3$, where n_2, n_3, p_2, p_3 are integers. Multiplying the resulting equation through by $(p_3 n_3)^4$ and introducing the new variable $x = p_3 n_3 x_1$, we obtain

$$x^4 + px^2 + n^2 = 0, \tag{1}$$

where $p = p_2 p_3 n_3^2$, $n = n_2 n_3 p_3^2$ are both integers.

The following theorem can be easily established.

THEOREM 1. *Necessary and sufficient conditions for the irreducibility, in the field of rational numbers, of the equation* (1) *with p and n rational, are that each of*

$$p^2 - 4n^2, \quad 2n - p, \quad -2n - p \tag{2}$$

be different from a perfect square in the field of rational numbers.

3. **Associativity conditions for a division algebra Γ of order sixteen.** Let us consider an equation (1) which is irreducible in the field of rational numbers and in which p and n are integers. It is proposed to develop algebras of type Γ, of order sixteen, on the basis of this equation (1).

Since this equation is irreducible in the field of rational numbers, $n \neq 0$; thus the roots of (1) are

$$x_1 = i, \quad x_2 = \theta_1(i) = -i, \quad x_3 = \theta_2(i) = \frac{i^3 + pi}{n} = -\frac{n}{i},$$

$$x_4 = \theta_3(i) \equiv \theta_1[\theta_2(i)] = \theta_2[\theta_1(i)] = \frac{-i^3 - pi}{n} = \frac{n}{i}.$$

* See R. J. Garver, *A normal form for certain quartics*, Messenger of Mathematics, vol. 56, No. 12, pp. 184–6.

Moreover, the Galois group of equation (1) for the field of rational numbers is

$$G_4 = \{I, \Theta_1, \Theta_2, \Theta_3 \equiv \Theta_1\Theta_2 = \Theta_2\Theta_1\},$$

where

$$\Theta_1 = (12)(34), \quad \Theta_2 = (14)(23).$$

In view of Dickson's memoir on *New division algebras*, §12, consider the quartic equation (1) irreducible in the field R of rational numbers. The roots $i, \theta_1(i), \theta_2(i), \theta_3(i)$ of equation (1) are rational functions of i with coefficients in R such that

$$\theta_1\theta_1 = \theta_2\theta_2 = \theta_3\theta_3 = i, \quad \theta_1\theta_2 = \theta_2\theta_1 = \theta_3, \quad \theta_1\theta_3 = \theta_3\theta_1 = \theta_2, \quad \theta_2\theta_3 = \theta_3\theta_2 = \theta_1,$$

where $\theta_r\theta_s$ denotes $\theta_r[\theta_s(i)]$. Under the law of multiplication

$$(a + bj_1)(c + dj_1) = ac + gbd(\theta_1) + [ad + bc(\theta_1)]j_1,$$

where a, b, c, d, g are in $R(i)$, the elements $a+bj_1$ form an associative algebra Σ if and only if g is in the field $R(i^2)$ (that is, if and only if $g=g(\theta_1)$), and they form an associative division algebra Σ if and only if g is in the field $R(i^2)$, but is not the norm, relative to $R(i^2)$, of any number of $R(i)$. For $A=a+bj_1$, write $A'=a(\theta_2)+b(\theta_2)\alpha j_1$, where α and γ (as defined in Dickson's paper) are in $R(i)$. Under the law of multiplication

$$(A_0 + A_1j_2)(B_0 + B_1j_2) = (A_0B_0 + A_1B_1'\gamma) + (A_0B_1 + A_1B_0')j_2,$$

where A_0, A_1, B_0, B_1 are in Σ, the elements $A_0+A_1j_2$ form an associative algebra Γ, when Σ is one, if and only if

$$\alpha\alpha(\theta_1)g = g(\theta_2), \qquad \alpha\alpha(\theta_2)\gamma(\theta_1) = \gamma = \gamma(\theta_2),$$

and they form an associative division algebra Γ, when Σ is one, if and only if

$$\gamma \neq X'X \text{ for any } X \text{ in } \Sigma,$$

and

$$\alpha\alpha(\theta_1)g = g(\theta_2), \qquad \alpha\alpha(\theta_2)\gamma(\theta_1) = \gamma = \gamma(\theta_2).$$

This Γ is an algebra over the field R with the sixteen basal units i^rj_s $(r, s=0, 1, 2, 3)$, where $j_3=j_1j_2$, $j_0=1$. On the other hand, the algebra Γ may be regarded as an algebra of order four over the field $R(i)$ with the basal units

$$1, j_1, j_2, j_3 = j_1j_2.$$

Special cases of the laws of multiplication give

$$j_1^2 = g, \; j_2^2 = \gamma, \; j_2j_1 = \alpha j_1j_2, \; j_ri = \theta_r(i)j_r, \; j_r\phi(i) = \phi[\theta_r(i)]j_r \; (r = 0,1,2,3).$$

Necessary and sufficient conditions that the algebra Γ be associative are the following:

$$j_1^2 = g(i) = g(\theta_1), \tag{3}$$

$$j_2^2 = \gamma(i) = \gamma(\theta_2), \tag{4}$$

$$\alpha\alpha(\theta_1)g = g(\theta_2), \tag{5}$$

$$\alpha\alpha(\theta_2) = \gamma(i)\gamma^{-1}(\theta_1), \tag{6}$$

where α, g, γ are rational functions of i with coefficients in R.

Being an even function of i, g must be of the form

$$g = ai^2 + b, \tag{7}$$

where a and b are both rational numbers. Also, since $\gamma = \gamma(\theta_2)$,

$$\gamma(i) = \gamma\left(-\frac{n}{i}\right);$$

whence, where c, d, e, f are all rational numbers,

$$\begin{aligned}\gamma &= ci^3 + di^2 + ei + f \\ &= i^3\left(\frac{e}{n} - \frac{cp}{n}\right) + i^2(-d) + i\left(cn - \frac{cp^2}{n} + \frac{ep}{n}\right) + (f - dp).\end{aligned}$$

Equating coefficients and combining results, we have

$$d = 0, \quad \frac{e}{n} = c\left(1 + \frac{p}{n}\right).$$

Therefore

$$\gamma = ci^3 + c(n + p)i + f. \tag{8}$$

Define a number η of $R(i)$ in terms of α, g, γ as follows:

$$j_3^2 = j_1j_2j_1j_2 = j_1(\alpha j_1)j_2j_2 = \alpha(\theta_1)j_1^2j_2^2 = \alpha(\theta_1)g\gamma = \eta. \tag{9}$$

By equations (5) and (6),

$$\alpha\alpha(\theta_1)g\gamma\gamma(\theta_1) = g(\theta_2)\gamma\gamma(\theta_1) = g(\theta_2)\alpha\alpha(\theta_2)\gamma(\theta_1)\gamma(\theta_1).$$

Hence, since α and γ are both different from zero,

$$\alpha(\theta_1)g\gamma = \alpha(\theta_2)g(\theta_2)\gamma(\theta_1).$$

Therefore, by definition of η, $\eta = \eta(\theta_1\theta_2)$. Moreover, by equation (5),

$$\eta\eta(\theta_1) = g\alpha\alpha(\theta_1)g\gamma\gamma(\theta_1) = gg(\theta_2)\gamma\gamma(\theta_1).$$

The associativity conditions (3)-(6) are now replaced by

$$j_1^2 = g = g(\theta_1), \tag{3'}$$

$$j_2^2 = \gamma = \gamma(\theta_2), \tag{4'}$$

$$j_3^2 = \eta = \eta(\theta_3), \tag{5'}$$

$$\eta\eta(\theta_1) = gg(\theta_2)\gamma\gamma(\theta_1). \tag{6'}$$

Evidently η is of the form

$$\eta = ri^3 + si^2 + ti + u,$$

where r, s, t, u are rational numbers. Therefore, in view of (5′),

$$ri^3 + si^2 + ti + u = i^3\left(\frac{rp}{n} - \frac{t}{n}\right) - si^2 + i\left(\frac{rp^2}{n} - \frac{tp}{n} - rn\right) + u - sp.$$

As a result of equating coefficients of like powers of i, we obtain

$$\eta = ri^3 + r(p - n)i + u. \tag{10}$$

Equation (6′) imposes a condition which becomes, on simplification,

$$r^2(n^2p - 2n^3) + u^2 = (a^2n^2 - abp + b^2)(c^2n^2p + 2c^2n^3 + f^2). \tag{11}$$

The problem of satisfying the associativity conditions (3′)-(6′) now becomes the problem of satisfying equation (11) with rational values for r, u, a, b, c, f.

4. **A type of diophantine equation.** When we write

$$h = n^2p - 2n^3, \quad U_1 = u, \quad C_1 = c,$$
$$l = n^2p + 2n^3, \quad A_1 = a, \quad F_1 = f,$$
$$R_1 = r, \quad B_1 = b,$$

equation (11) becomes

$$hR_1^2 + U_1^2 = (n^2A_1^2 - pA_1B_1 + B_1^2)(lC_1^2 + F_1^2), \tag{12}$$

where n and p are regarded as parameters and the capital letters as variables. It will now be proved that all *rational* solutions of equation (12) are known when all *integral* solutions are known.

Consider any rational solution:

$$R_1 = R/D,\ U_1 = U/D,\ A_1 = A/D,\ B_1 = B/D,\ C_1 = C/D,\ F_1 = F/D, \tag{13}$$

where R, U, A, B, C, F, D are integers. Therefore, placing these values in equation (12) and multiplying through by D^4, we obtain

$$h(DR)^2 + (DU)^2 = (n^2A^2 - pAB + B^2)(lC^2 + F^2).$$

Therefore,

$$R_1 = DR,\ U_1 = DU,\ A_1 = A,\ B_1 = B,\ C_1 = C,\ F_1 = F \tag{14}$$

is an integral solution of (12). Consequently, any rational solution (13) is obtained from an appropriate integral solution (14) by dividing the first two values in (14) by D^2, and the last four by D.

Multiply equation (12) through by 4 and rewrite it in the form

$$h(2R_1)^2 + (2U_1)^2 = \{n^2(2A_1)^2 - p(2A_1)(2B_1) + (2B_1)^2\}\{lC_1^2 + F_1^2\}. \tag{15}$$

Make now the following substitutions in (15):

$$A_1 = B_0, \qquad 2R_1 = R_2,$$
$$2B_1 = -A_0 + pB_0, \qquad 2U_1 = U, \tag{16}$$
$$p^2 - 4n^2 = -m.$$

Equation (15) then becomes

$$hR_2^2 + U^2 = (A_0^2 + mB_0^2)(lC_1^2 + F_1^2). \tag{17}$$

All integral solutions of (12) give rise, by virtue of relations (16), to integral solutions of (17). Moreover, the totality of those integral solutions of (17) for which are satisfied the following conditions (18), give rise to all the integral solutions of (12):

$$R_2 \equiv 0,\ U \equiv 0,\ A_0 \equiv pB_0 \pmod 2. \tag{18}$$

Placing

$$a_1 = 2n - p,\ b_1 = -2n - p,\ R_0 = nR_2,\ C_0 = nC_1, \tag{19}$$

we bring equation (17) to the form

$$U^2 - a_1R_0^2 = (A_0^2 - a_1b_1B_0^2)(F_1^2 - b_1C_0^2). \tag{20}$$

All integral solutions of (12) are obtained by securing all integral solutions of (20) for which the following conditions (21) are satisfied:

$$\begin{aligned} R_0 &\equiv 0 \pmod{2n}, & U &\equiv 0 \pmod 2, \\ C_0 &\equiv 0 \pmod n, & A_0 &\equiv pB_0 \pmod 2. \end{aligned} \tag{21}$$

The problem is now reduced to that of considering the solution* in integers (not all zero) of equation (20), where, in view of conditions (2), each of a_1, b_1, a_1b_1 is different from a perfect square.

5. **Modification of the problem.** In the Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 464–7, Dickson derived† the result that all integral solutions of the equation

$$x^2 - my^2 = zw$$

are given by

$$\begin{aligned} z &= s(el^2 + 2flq + gq^2), & w &= s(en^2 - 2fnl + gl^2), \\ x &= \pm s(eln + fnq - flt - gql), & y &= s(lt + nq), \end{aligned} \tag{22}$$

* This problem in diophantine analysis cannot be solved by the known methods of composition of forms. For, to obtain solutions by composition, the determinants of $U^2 - a_1R_0^2$, $A^2 - a_1b_1B_0^2$, and $F^2 - b_1C_0^2$ must be in the ratios of three integral squares (cf. C. F. Gauss, *Disquisitiones Arithmeticae*, art. 235). That is, a_1b_1/a_1 = perfect square. This is contrary to restrictions (2) placed on b_1.

† See also a paper by Dickson in the same Bulletin, vol. 27 (1920–21), pp. 353–365. In that paper Dickson was led to the present theorem by the theory of ideals.

where s ranges over the set of rational integers and where e, f, g take only those sets of integral values (finite in number) for which the form $el^2+2flq+gq^2$ is a reduced* quadratic form having the same discriminant $4m$ as x^2-my^2, so that

$$f^2 - eg = m.$$

Consider equation (20) in which it is assumed that each of the integers a_1, b_1, a_1b_1 is different from a perfect square. Let a_0^2, b_0^2 be the greatest integral squares contained in a_1, b_1, and let c be the greatest common divisor of a_1/a_0^2, b_1/b_0^2. Write

$$a_1 = aca_0^2, \qquad b_1 = bcb_0^2.$$

Consequently, we can state the restrictions on a, b, c as follows:

(23) c is positive; no one of a, b, c possesses a square factor (other than 1); no two of a, b, c have a common factor greater than 1; and, moreover, whenever $c=1$, both a and b are different from 1.

Equation (20) can be written as

$$(20') \qquad U^2 - ac(a_0R_0)^2 = \{A_0^2 - ab(ca_0b_0B_0)^2\}\{F_1^2 - bc(b_0C_0)^2\}.$$

Place

$$(24) \qquad R = a_0R_0,\ A = A_0,\ B = ca_0b_0B_0,\ F = F_1,\ C = b_0C_0.$$

Equation (20′) now becomes

$$(25) \qquad U^2 - acR^2 = (A^2 - abB^2)(F^2 - bcC^2),$$

where a, b, c are subject to conditions (23).

All integral solutions of equation (20) are obtained from all those integral solutions of equation (25) which satisfy conditions (26) following:

$$(26) \qquad R \equiv 0 \pmod{a_0},\ B \equiv 0 \pmod{ca_0b_0},\ C \equiv 0 \pmod{b_0}.$$

Let $U=u$, $R=r$, $A=\bar{a}$, $B=\bar{b}$, $F=\bar{f}$, $C=\bar{c}$ be an integral solution of (25). Then $U=ua_0^2b_0^2c$, $R=ra_0^2b_0^2c$, $A=\bar{a}a_0^2b_0c$, $B=\bar{b}a_0^2b_0c$, $F=\bar{f}b_0$, $C=\bar{c}b_0$ is a solution of (25) satisfying conditions (26). Therefore, equations (20) and (25) are *simultaneously* solvable (in integers, not all zero), or not solvable. The same is true also of equations (12) and (20), and, consequently, of equations (12) and (25).

Placing

$$Z = A^2 - abB^2, \quad W = F^2 - bcC^2,$$

* That is, a single form—in fact, not necessarily reduced—is employed from each class of quadratic forms of discriminant $4m$.

we may write equation (25) as

$$U^2 - acR^2 = ZW\ ;$$

hence, in view of Dickson's solution of the equation $x^2-my^2=zw$, it is seen that solving equation (25) in integers is equivalent to solving in integers

$$A^2 - abB^2 = s(eL_1^2 + 2fL_1Q_1 + gQ_1^2), \tag{27a}$$

$$F^2 - bcC^2 = s(eN_1^2 - 2fN_1T_1 + gT_1^2). \tag{27b}$$

After multiplying equations (27a), (27b) through by e, making the transformation

$$L = eL_1 + fQ_1,\ N = eN_1 - fT_1,\ Q = Q_1,\ T = T_1, \tag{28}$$

and replacing f^2-eg by its value ac, we have equations (27a), (27b) replaced by

$$eA^2 - eabB^2 - sL^2 + sacQ^2 = 0, \tag{29a}$$

$$eF^2 - ebcC^2 - sN^2 + sacT^2 = 0. \tag{29b}$$

All integral solutions of (27a), (27b) are obtained from all those integral solutions of (29a), (29b) such that the following conditions hold:

$$L \equiv fQ\ ;\ N \equiv -fT \pmod{e}. \tag{30}$$

Since equation (25) and equations (29a), (29b) are simultaneously solvable, or not solvable, it is now our purpose to obtain necessary and sufficient conditions for the solvability in integers of equations (29a), (29b).

6. **Meyer's solution of $ax^2+by^2+cz^2+du^2=0$.** A set of necessary and sufficient conditions for the solvability of the equation

$$ax^2 + by^2 + cz^2 + du^2 = 0 \tag{31}$$

in integers (not all zero), where a, b, c, d are all non-zero integers, has been obtained by A. Meyer.*

It is assumed that a, b, c, d are not only non-zero but also without square factors, and that they are such that no three have a common factor (greater than 1). The greatest common divisor of a and b is designated by (a, b). Let

$$a = (a,b)(a,c)(a,d)\alpha, \qquad b = (b,a)(b,c)(b,d)\beta,$$

$$c = (c,a)(c,b)(c,d)\gamma, \qquad d = (d,a)(d,b)(d,c)\delta.$$

---

* Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, vol. 29 (1884), pp. 209–222. Cf. also P. Bachmann, *Zahlentheorie*, Part IV, *Die Arithmetik der Quadratischen Formen*, I, pp. 259–266.

Necessary conditions for the solvability of equation (31) in integers not all zero are the following:

I: a, b, c, d are not all of the same sign.

II: $-(a, c)(a, d)(b, c)(b, d)\gamma\delta$ is a quadratic residue of (a, b);
$-(a, b)(a, d)(c, b)(c, d)\beta\delta$ " " " " " (a, c);
$-(a, b)(a, c)(d, b)(d, c)\beta\gamma$ " " " " " (a, d);
$-(b, a)(b, d)(c, a)(c, d)\alpha\delta$ " " " " " (b, c);
$-(b, a)(b, c)(d, a)(d, c)\alpha\gamma$ " " " " " (b, d);
$-(c, a)(c, b)(d, a)(d, b)\alpha\beta$ " " " " " (c, d).

Necessary and sufficient conditions are: Conditions I and II above, and

III: either

(i) $abcd \equiv 2, 3, 5, 6, 7 \pmod 8$; or

(ii) $abcd \equiv 1$ and $a+b+c+d \equiv 0 \pmod 8$; or

(iii) $abcd \equiv 4 \pmod 8$, and, if a and b are even and c and d odd, either $\frac{1}{4}abcd \equiv 3, 5, 7 \pmod 8$, or $\frac{1}{4}abcd \equiv 1 \pmod 8$ and $a/2+b/2+c+d \equiv \{(cd)^2-1\}/2 \pmod 8$.

7. **Necessary and sufficient conditions for the solvability of the diophantine equation.** Necessary and sufficient conditions will here be derived for the solvability in integers of equations (29a), (29b), and hence for the solvability of the original diophantine equation (20).

Now, $e \neq 0$ since ac is not a perfect square; moreover, it is initially assumed that $s \neq 0$ since in the contrary case the problem becomes a trivial one. Where ϵ^2, σ^2 are the greatest integral squares contained in e, s, respectively, write

$$\begin{aligned} e &= e_1\epsilon^2, & s &= s_1\sigma^2, \\ e_1 &= (e_1, s_1)e_2, & s_1 &= (e_1, s_1)s_2, \\ e_2 &= (e_2, a)(e_2, b)e_3, & ab &= (e_2, a)(e_2, b)a_2b_2, \\ s_2 &= (s_2, a)(s_2, c)s_3, & ac &= (s_2, a)(s_2, c)a_3c_1. \end{aligned}$$

Noting that $(e_3, c_1) = (e_3, c) = (e_2, c)$ and $(b_2, s_3) = (b_2, s_2) = (b, s_2)$, we write

$$e_3 = (e_2, c)e_4, \quad c_1 = (e_2, c)c_2, \quad b_2 = (b, s_2)b_3, \quad s_3 = (b, s_2)s_4.$$

Moreover, place

$$\begin{aligned} A_2 &= \epsilon A, & B_3 &= \epsilon B, & L_2 &= \sigma L, \\ Q_2 &= \sigma Q, & B_4 &= (e_2, a)(e_2, b)B_3, & Q_3 &= (s_2, a)(s_2, c)Q_2, \\ & & L_2 &= (e_2, c)L_3, & A_2 &= (b, s_2)A_3. \end{aligned}$$

Equation (29a) then becomes

$$(e_2,a)(e_2,b)(b,s_2)e_4A_3^2 - e_4a_2b_3B_4^2 - (s_2,a)(s_2,c)(e_2,c)s_4L_3^2 + s_4a_3c_2Q_3^2 = 0. \tag{29a'}$$

Let $e_1, s_1, e_2, s_2, s_3, a_3, c_1$ be defined as above, and write

$$e_2 = (e_2,b)(e_2,c)e_5, \quad bc = (e_2,b)(e_2,c)b_2c_3,$$

where $c_3=(s_2, c)c_2$. Moreover, since $(e_5, a_3)=(e_5, a)=(e_2, a_3)=(e_2, a)$ and since $(e_5, c_1)=(e_5, c)=1$, write

$$e_5 = (e_2,a)e_4, \quad a_3 = (e_2,a)a_4, \quad b_2 = (b,s_2)b_3, \quad s_3 = (b,s_2)s_4,$$

where $a_2=(s_2,a)a_4$. Also place

$$F_2 = \epsilon F, \quad C_3 = \epsilon C, \quad N_2 = \sigma N,$$
$$T_2 = \sigma T, \quad C_4 = (e_2,b)(e_2,c)C_3, \quad T_3 = (s_2,a)(s_2,c)T_2,$$
$$N_2 = (e_2,a)N_3, \quad F_2 = (b,s_2)F_3.$$

Equation (29b) then becomes

$$(e_2,b)(e_2,c)(b,s_2)e_4F_3^2 - e_4b_3c_3C_4^2 - (s_2,a)(s_2,c)(e_2,a)s_4N_3^2 + s_4a_4c_1T_3^2 = 0. \tag{29b'}$$

All integral solutions of equations (29a), (29b) are obtained by securing all integral solutions $A_3, B_4, L_3, Q_3; F_3, C_4, N_3, T_3$ of equations (29a'), (29b'), respectively, such that

$$\begin{aligned} (b,s_2)A_3 &\equiv 0 \pmod{\epsilon}, & B_4 &\equiv 0 \pmod{(e_2,a)(e_2,b)\epsilon}, \\ (e_2,c)L_3 &\equiv 0 \pmod{\sigma}, & Q_3 &\equiv 0 \pmod{(s_2,a)(s_2,c)\sigma}, \\ (b,s_2)F_3 &\equiv 0 \pmod{\epsilon}, & C_4 &\equiv 0 \pmod{(e_2,b)(e_2,c)\epsilon}, \\ (e_2,a)N_3 &\equiv 0 \pmod{\sigma}, & T_3 &\equiv 0 \pmod{(s_2,a)(s_2,c)\sigma}. \end{aligned} \tag{32}$$

Meyer's *preliminary conditions* are satisfied by both equations (29a'), (29b'). In the notation of Meyer, for equations (29a'), (29b'), respectively, we have the following two sets of four equations each:

$$\begin{aligned} (e_2,b)(b,s_2)e_4 &= |e_4|\alpha, \\ -e_4a_2b_3 &= |e_4|(s_2,a)|a_4|\beta, \\ -(s_2,c)(e_2,c)s_4 &= |s_4|\gamma, \\ s_4a_3c_2 &= (e_2,a)|a_4|\cdot|s_4|\delta; \\ (e_2,b)(b,s_2)e_4 &= |e_4|\alpha_1, \\ -e_4b_3c_3 &= |e_4|(s_2,c)c_2\beta_1, \\ -(s_2,a)(e_2,a)s_4 &= |s_4|\gamma_1, \\ s_4a_4c_1 &= (e_2,c)c_2|s_4|\delta_1. \end{aligned}$$

CONDITION* I. Either s and e are of the same sign; or s and e are of opposite sign, and either a is positive and b negative, or a and b are both positive.

CONDITION II.

(33a)	ac	is a quadratic residue of	$\lvert e_4 \rvert$.
(33b)	$-(s_2, c)e_3s_4a_4b_3$	" " " " "	(e_2, a).
(33c)	$-(s_2, a)e_5s_4b_3c_2$	" " " " "	(e_2, c).
(33d)	$-(e_2, b)e_4s_3a_4c_2$	" " " " "	(s_2, a) (s_2, c).
(33e)	e_2s_2	" " " " "	$\lvert a_4 \rvert c_2$.
(33f)	ab	" " " " "	$\lvert s_4 \rvert$.
(33g)	bc	" " " " "	$\lvert s_4 \rvert$.

Conditions (33a)-(33e) are equivalent, respectively, to the following conditions (33a')-(33e'):

(33a') For every (positive) odd prime factor of $\lvert e_4 \rvert$, a and c are simultaneously quadratic residues or simultaneously quadratic non-residues of the chosen odd prime factor of $\lvert e_4 \rvert$.

(33b') For every odd prime factor of (e_2, a), exactly an even number (including zero) of $-(s_2, c)$, e_3, s_4, a_4, b_3 are quadratic non-residues of the chosen odd prime factor of (e_2, a).

(33c') For every odd prime factor of (e_2, c), exactly an even number (including zero) of $-(s_2, a)$, e_5, s_4, b_3, c_2 are quadratic non-residues of the chosen odd prime factor of (e_2, c).

(33d') For every odd prime factor of (s_2, ac), exactly an even number (including zero) of $-(e_2, b)$, e_4, s_3, a_4, c_2 are quadratic non-residues of the chosen odd prime factor of (s_2, ac).

(33e') For every odd prime factor of $\lvert a_4 \rvert c_2$, e_2 and s_2 are simultaneously quadratic residues or simultaneously quadratic non-residues of the chosen odd prime factor of $\lvert a_4 \rvert c_2$.

Let s_j $(j=6, 7, \cdots, \mu+5)$ be the distinct (positive) odd prime factors of $\lvert s_4 \rvert$, and let a_i $(i=5, 6, \cdots, \lambda+4)$, b_i $(i=5, 6, \cdots, \rho+4)$, and c_i $(i=5, 6, \cdots, \nu+4)$ be the distinct (positive) odd prime factors of a, b, c, respectively.

* We lay aside the trivial case in which one (or both) of equations (29a'), (29b') has merely the zero solution.

Conditions (33f), (33g) can be combined and replaced by the equivalent statement that, for every positive odd prime factor of $|s_4|$, each of a, b, c is simultaneously a quadratic residue or simultaneously a quadratic non-residue of the chosen odd prime factor of $|s_4|$. This condition will be given in another form. The condition (on $|s_4|$) that a, b, c be simultaneously quadratic residues of a chosen s_j, is treated first, under four separate cases. Similarly, the alternative condition, that a, b, c be simultaneously quadratic non-residues of a chosen s_j, is treated later.

Case 1A. $a=2a_5a_6\cdots a_{\lambda+4}$.

Now a is a quadratic residue of $|s_4|$ if and only if each of the Legendre symbols

$$\left(\frac{a}{s_6}\right),\left(\frac{a}{s_7}\right),\cdots,\left(\frac{a}{s_{\mu+5}}\right)$$

has the value $+1$. Now

$$\left(\frac{a}{s_j}\right)=\left(\frac{2}{s_j}\right)\left(\frac{a_5}{s_j}\right)\left(\frac{a_6}{s_j}\right)\cdots\left(\frac{a_{\lambda+4}}{s_j}\right)$$

$$=(-1)^{\zeta}\left(\frac{s_j}{a_5}\right)\left(\frac{s_j}{a_6}\right)\cdots\left(\frac{s_j}{a_{\lambda+4}}\right),$$

where $\zeta=(s_j^2-1)/8+\frac{1}{4}(s_j-1)\{\sum_{i=5}^{\lambda+4}a_i-\lambda\}$. Hence

(34a) either $s_j^2+2(s_j-1)\{\sum_{i=5}^{\lambda+4}a_i-\lambda\}\equiv 1 \pmod{16}$, and s_j is a quadratic non-residue of exactly an even number (including zero) of the prime factors of $a/2$; or $s_j^2+2(s_j-1)\{\sum_{i=5}^{\lambda+4}a_i-\lambda\}\equiv 9 \pmod{16}$, and s_j is a quadratic non-residue of exactly an odd number of the prime factors of $a/2$.

Case 1B. $a=-2a_5a_6\cdots a_{\lambda+4}$.

In this case the condition, designated by (34b), is the same as condition (34a) when λ of (34a) is replaced (except in the upper limit of summation) by $\lambda-2$ and a by $-a$.

Case 1C. $a=a_5a_6\cdots a_{\lambda+4}$.

(34c) Either $(s_j-1)\{\sum_{i=5}^{\lambda+4}a_i-\lambda\}\equiv 0 \pmod 8$, and s_j is a quadratic non-residue of exactly an even number (including zero) of the prime factors of a; or $(s_j-1)\{\sum_{i=5}^{\lambda+4}a_i-\lambda\}\equiv 4 \pmod 8$, and s_j is a quadratic non-residue of exactly an odd number of the prime factors of a.

Case 1D. $a=-a_5a_6\cdots a_{\lambda+4}$.

The condition in this case, designated by (34d), is the same as (34c) after λ has been replaced (except in the upper limit of summation) by $\lambda-2$ and a by $-a$.

The corresponding conditions for b, denoted by (35a), (35b), (35c), (35d), are obtained from conditions (34a), (34b), (34c), (34d), respectively, by replacing a by b, λ by ρ, a_i by b_i.

The corresponding conditions for c, denoted by (36a), (36c), are obtained from conditions (34a), (34c), respectively, by replacing a by c, λ by ν, a_i by c_i.

The alternative condition, that a, b, c are simultaneously quadratic non-residues of a chosen s_j, will now be given in more suitable form.

Case 2A. $a=2a_5a_6\cdots a_{\lambda+4}$.

(37a) Either $s_j^2+2(s_j-1)\{\sum_{i=5}^{\lambda+4}a_i-\lambda\}\equiv 1 \pmod{16}$, and s_j is a quadratic non-residue of exactly an odd number of the prime factors of $a/2$; or $s_j^2+2(s_j-1)\{\sum_{i=5}^{\lambda+4}a_i-\lambda\}\equiv 9 \pmod{16}$, and s_j is a quadratic non-residue of exactly an even number (including zero) of the prime factors of $a/2$.

Case 2B. $a=-2a_5a_6\cdots a_{\lambda+4}$.

In formulating the condition, designated by (37b), for this case, modify the statement of condition (34b) for Case 1B in the same way as the statement of condition (34a) is modified to give condition (37a) for Case 2A.

Case 2C. $a=a_5a_6\cdots a_{\lambda+4}$.

The condition here, designated by (37c), is an analogous modification of condition (34c) for Case 1C.

Case 2D. $a=-a_5a_6\cdots a_{\lambda+4}$.

Similarly, the condition, designated by (37d), for this case is again an analogous modification of condition (34d) for Case 1D.

The corresponding set of conditions for b, denoted by (38a), (38b), (38c), (38d), are obtained from conditions (37a), (37b), (37c), (37d), respectively, by replacing a by b, λ by ρ, a_i by b_i.

The corresponding set of conditions for c, denoted by (39a), (39c), are obtained from conditions (37a), (37c), respectively, by replacing a by c, λ by ν, a_i by c_i.

All the foregoing conditions,—namely, (34a), $\cdots$, (34d), (35a), $\cdots$, (35d), (36a), (36c), (37a), $\cdots$, (37d), (38a), $\cdots$, (38d), (39a), (39c)—are stated for any arbitrarily chosen s_j of the μ distinct positive odd prime factors of $|s_4|$; hence, it is to be understood that in Condition II $j=6, 7, \cdots, \mu+5$.

Define

$$h=(e_2,a)(e_2,b)(b,s_2)e_4-e_4a_2b_3-(s_2,a)(s_2,c)(e_2,c)s_4+s_4a_3c_2,$$

$$k=\tfrac{1}{2}\{(e_2,a)(e_2,b)(b,s_2)e_4-e_4a_2b_3\}-(s_2,a)(s_2,c)(e_2,c)s_4+s_4a_3c_2,$$

$$p=(e_2,b)(e_2,c)(b,s_2)e_4-e_4b_3c_3-(s_2,a)(s_2,c)(e_2,a)s_4+s_4a_4c_1,$$

$$q=\tfrac{1}{2}\{(e_2,b)(e_2,c)(b,s_2)e_4-e_4b_3c_3\}-(s_2,a)(s_2,c)(e_2,a)s_4+s_4a_4c_1.$$

CONDITION IIIA. Either

(i) $e_4^2 s_4^2 a^2 bc \equiv 2, 3, 5, 6, 7 \pmod 8$; or

(ii) $e_4^2 s_4^2 a^2 bc \equiv 1$ and $h \equiv 0 \pmod 8$; or

(iii) $e_4^2 s_4^2 a^2 bc \equiv 4 \pmod 8$, and, if e_4 is even, either $\frac{1}{4} e_4^2 s_4^2 a^2 bc \equiv 3, 5, 7 \pmod 8$, or $\frac{1}{4} e_4^2 s_4^2 a^2 bc \equiv 1 \pmod 8$ and $k \equiv \frac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8$.

CONDITION IIIB. Either

(i) $e_4^2 s_4^2 abc^2 \equiv 2, 3, 5, 6, 7 \pmod 8$; or

(ii) $e_4^2 s_4^2 abc^2 \equiv 1$ and $p \equiv 0 \pmod 8$; or

(iii) $e_4^2 s_4^2 abc^2 \equiv 4 \pmod 8$, and, if e_4 is even, either $\frac{1}{4} e_4^2 s_4^2 abc^2 \equiv 3, 5, 7 \pmod 8$, or $\frac{1}{4} e_4^2 s_4^2 abc^2 \equiv 1 \pmod 8$ and $q \equiv \frac{1}{2}\{(acs_4^2)^2 - 1\}$.

The results may now be summarized. Necessary and sufficient conditions for the solvability in integers of equations (29a′), (29b′), when it is assumed that $s \neq 0$, are given as follows:

Condition I and conditions (33a)-(33e), or the equivalent conditions (33a′)-(33e′), must hold; and, according to the character of a, b, c, for every odd prime factor of $|s_4|$ *either* one condition must hold from each of the three sets of conditions

(34a), $\cdots$, (34d); (35a), $\cdots$, (35d); (36a), (36c),

or one condition must hold from each of the three sets of conditions

(37a), $\cdots$, (37d); (38a), $\cdots$,(38d); (39a), (39c).

Conditions IIIA, IIIB must hold.

There are combinations among these last conditions (IIIA, IIIB) which are not consistent.

There are twenty-five possible ways of combining a condition of IIIA (i) for equation (29a′) with a condition of IIIB (i) for equation (29b′); but it is readily seen that the following twelve combinations are inconsistent:

$$e_4^2 s_4^2 a^2 bc \equiv 2,6, \quad e_4^2 s_4^2 abc^2 \equiv 3,5,7 \pmod 8;$$

$$e_4^2 s_4^2 a^2 bc \equiv 3,5,7, \quad e_4^2 s_4^2 abc^2 \equiv 2,6 \pmod 8.$$

By taking $e = s = 1$, we find that the remaining thirteen pairs of congruences can be satisfied.

Of the five ways of combining a condition of IIIA (i) with the condition IIIB (ii), it is seen that three give consistent combinations and the remaining two give inconsistent combinations.

Of the twenty-five ways of combining a condition of IIIA (i) with a condi-

tion of IIIB (iii), it is seen that only two give consistent combinations: namely,

$$e_4^3 s_4^3 a^2bc \equiv 2,6, \quad e_4^3 s_4^3 abc^2 \equiv 4 \pmod 8,$$

where e_4 is odd. These two pairs of congruences can be satisfied by $e=s=1$, $a=1$, $b=17$, $c=2$; $e=s=1$, $a=1$, $b=3$, $c=2$, respectively.

Of the five ways of combining the condition IIIA (ii) with a condition of IIIB (i), three give consistent combinations and two give inconsistent combinations; the combination of condition IIIA (ii) with condition IIIB (ii) is a consistent one; of the five ways of combining the condition IIIA (ii) with a condition of IIIB (iii), all five give inconsistent combinations.

Of the twenty-five ways of combining a condition of IIIA (iii) with a condition of IIIB (i), it is seen that only two give consistent combinations: namely,

$$e_4^3 s_4^3 a^2bc \equiv 4, \quad e_4^3 s_4^3 abc^2 \equiv 2,6 \pmod 8,$$

where e_4 is odd. These two pairs of congruences can be satisfied by $e=s=1$, $a=2$, $b=17$, $c=1$; $e=s=1$, $a=2$, $b=3$, $c=1$, respectively. All five ways of combining a condition of IIIA (iii) with condition IIIB (ii) give inconsistent combinations.

Of the twenty-five ways of combining a condition of IIIA (iii) with a condition of IIIB (iii), it is found that fifteen give consistent combinations and ten give inconsistent combinations. Here it is somewhat more difficult to determine whether or not a combination is consistent. The following fifteen combinations are the only ones which can be satisfied.

The combination

$$e_4^2 s_4^2 a^2bc \equiv 4, \quad e_4^2 s_4^2 abc^2 \equiv 4 \pmod 8,$$

where e_4 is odd, can be satisfied by $e=1$, $s=2$, $a=7$, $b=5$, $c=1$.

The combination of

$$e_4^2 s_4^2 a^2bc \equiv 4, \quad \tfrac{1}{4} e_4^2 s_4^2 a^2bc \equiv x \pmod 8$$

with

$$e_4^2 s_4^2 abc^2 \equiv 4, \quad \tfrac{1}{4} e_4^2 s_4^2 abc^2 \equiv y \pmod 8,$$

where e_4 is even, is satisfied

by $e=2$, $s=1$, $a=7$, $b=5$, $c=23$, when $x=y=3$;
by $e=2$, $s=1$, $a=7$, $b=3$, $c=1$, when $x=3$, $y=5$;
by $e=14$, $s=a=1$, $b=7$, $c=5$, when $x=3$, $y=7$;
by $e=2$, $s=a=1$, $b=3$, $c=7$, when $x=5$, $y=3$;
by $e=2$, $s=a=1$, $b=5$, $c=17$, when $x=5$, $y=5$;

by $e=14$, $s=a=1$, $b=15$, $c=11$, when $x=5$, $y=7$;
by $e=14$, $s=1$, $a=5$, $b=7$, $c=1$, when $x=7$, $y=3$;
by $e=2$, $s=-1$, $a=11$, $b=7$, $c=1$, when $x=7$, $y=5$;
by $e=2$, $s=a=1$, $b=7$, $c=17$, when $x=7$, $y=7$.

The preceding combination with $y=1$, together with

$$q \equiv \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8,$$

is satisfied by $e=14$, $s=a=1$, $b=17$, $c=11$, when $x=3$; by $e=2$, $s=a=1$, $b=17$, $c=7$, when $x=7$. Moreover, the preceding combination with $x=1$, together with

$$k \equiv \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8$$

instead of

$$q \equiv \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8,$$

is satisfied by $e=14$, $s=1$, $a=11$, $b=17$, $c=1$, when $y=3$; by $e=2$, $s=1$, $a=7$, $b=17$, $c=1$, when $y=7$. The preceding combination with $x=y=1$, together with

$$k \equiv \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8$$

and

$$q \equiv \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8,$$

is satisfied by $e=2$, $s=1$, $a=-1$, $b=15$, $c=7$.

Since e_4 cannot at the same time be both odd and even, eight combinations are immediately excluded as inconsistent. As is later proved in Theorem 4 of §8, the remaining two ways of combining a condition of IIIA (iii) with a condition of IIIB (iii) give inconsistent combinations.

Finally, out of the 121 combinations of a condition of IIIA for equation (29a′) with a condition of IIIB for equation (29b′), there are exactly thirty-nine consistent combinations and eighty-two inconsistent combinations.

8. **Incompatibility of certain congruences.** It will here be shown, in Theorem 4, that, as stated in the preceding section, certain combinations of a condition from IIIA (iii) with a condition from IIIB (iii) are inconsistent. Theorems 2 and 3 are lemmas for Theorem 4. Under the (additional) restriction that e_2s_2 is relatively prime to abc, Theorems 5, 6, and 7 can be proved. In this case, there are exactly thirty-seven consistent combinations of a condition of IIIA for equation (29a′) with a condition of IIIB for equation (29b′).

Throughout the present section, let e, e_i $(i=1, 2, \cdots, 5)$, s, s_i $(i=1, 2, \cdots, 4)$, a, b, c, p, and q be integers defined as in §7.

THEOREM 2. *If* $e_4=2e_7$, $a\equiv b \pmod 8$, *and* $c\equiv 5a \pmod 8$, *then* $q\equiv 0 \pmod 8$.

By hypothesis,

$$a \equiv (e_2,b)(b,s_2)b_3 \pmod 8,$$

$$c \equiv (e_2,c)(s_2,c)c_2 \equiv 5a \pmod 8.$$

Then

$$a \equiv (e_2,b)(b,s_2)b_3 \equiv 5(e_2,c)(s_2,c)c_2 \pmod 8.$$

Since the square of an odd integer is congruent to 1 modulo 8, on multiplying both sides of the latter congruence by $(e_2, c)e_7b_3$ we obtain

$$(e_2,b)(e_2,c)(b,s_2)e_7 \equiv 5(s_2,c)c_2e_7b_3 \pmod 8$$

$$\equiv 5e_7b_3c_3 \pmod 8.$$

Similarly, since $ac \equiv 5 \pmod 8$,

$$(s_2,a)(s_2,c)(e_2,a)(e_2,c)a_4c_2 \equiv 5 \pmod 8;$$

moreover, multiplying both sides of this congruence by $(e_2, c)s_4a_4c_2$, we obtain

$$(s_2,a)(s_2,c)(e_2,a)s_4 \equiv 5(e_2,c)s_4a_4c_2 \pmod 8$$

$$\equiv 5s_4a_4c_1 \pmod 8.$$

Hence,

$$(e_2,b)(e_2,c)(b,s_2)e_7 - (s_2,a)(s_2,c)(e_2,a)s_4 \equiv 5(e_7b_3c_3 - s_4a_4c_1) \pmod 8;$$

and, since e_4 is relatively prime to a, b, c and e_7 is odd,

$$q \equiv 4(e_7b_3c_3 - s_4a_4c_1) \pmod 8$$

$$\equiv 0 \pmod 8.$$

THEOREM 3. *If* $e_4 = 2e_7$, $b \equiv c \pmod 8$, *and* $a \equiv 5b \pmod 8$, *then* $k \equiv 0 \pmod 8$.

The proof of this theorem is similar to that of the preceding theorem.

THEOREM 4. *If* $e_4 = 2e_7$, *the set of congruences*

$$e_4^2 s_4^2 a^2bc \equiv 4 \pmod 8, \qquad e_4^2 s_4^2 abc^2 \equiv 4 \pmod 8,$$

$$\tfrac{1}{4}e_4^2 s_4^2 a^2bc \equiv x \pmod 8, \qquad \tfrac{1}{4}e_4^2 s_4^2 abc^2 \equiv y \pmod 8$$

is inconsistent with the congruence $q \equiv \frac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8$ *when* $x = 5, y = 1$, *and is inconsistent with the congruence* $k \equiv \frac{1}{2}\{(acs_4^2)^2 - 1\} \pmod 8$ *when* $x = 1$, $y = 5$.

Since the square of an odd integer is congruent to 1 modulo 8, it is seen that, when $x = 5$, $y = 1$,

$$bc \equiv 5, \qquad ab \equiv 1 \pmod 8.$$

Therefore

$$a \equiv b, \qquad c \equiv 5b \equiv 5a, \qquad \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \equiv 4 \pmod 8.$$

Then

$$q \equiv \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \equiv 4 \pmod 8.$$

By Theorem 2 it is seen that $q \equiv 0 \pmod 8$. Hence Theorem 4 is proved when $x=5$, $y=1$.

If $x=1$, $y=5$,

$$b \equiv c, \qquad a \equiv 5b, \qquad \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \equiv 4 \pmod 8.$$

Then

$$k \equiv \tfrac{1}{2}\{(acs_4^2)^2 - 1\} \equiv 4 \pmod 8.$$

By Theorem 3 it is seen that $k \equiv 0 \pmod 8$. Thus Theorem 4 is proved when $x=1$, $y=5$.

The following theorems are of interest when it is assumed that $e_2 s_2$ is relatively prime to abc.

THEOREM 5. *If $e_4 = 2e_7$ and if $e_2 s_2$ is relatively prime to abc, then a necessary condition that each of e, e_i $(i=1, \cdots, 5)$, s, s_i $(i=1, \cdots, 4)$, a, b, c, as defined in §7, satisfy conditions* (23), (33a), (33e), (33f), (33g), *is that*

$$\zeta = \frac{a^2+c^2-2}{8} + \frac{|s_2|-1}{2}\cdot\frac{a-1}{2} + \frac{c-1}{2}\cdot\frac{s_2-1}{2} + \frac{e_7-1}{2}\cdot\frac{c-1}{2} + \frac{a-1}{2}\cdot\frac{|e_7|-1}{2} + \xi + \eta$$

be even, when we define

$$\xi = 0 \text{ if } s \text{ is positive}, \quad = \{|a|-1\}/2 \text{ if } s \text{ is negative};$$
$$\eta = 0 \text{ if } e \text{ is positive}, \quad = \{|a|-1\}/2 \text{ if } e \text{ is negative}.$$

THEOREM 6. *If the integers e, e_i $(i=1, \cdots, 5)$, s, s_i $(i=1, \cdots, 4)$, a, b, c, as defined in §7, satisfy conditions* (23), (33a), (33e), (33f), (33g), *and condition* I *of §7, and if, when $e_4 = 2e_7$, $e_2 s_2$ is relatively prime to abc, then the congruences*

$$c \equiv 5a \pmod 8, \quad a^2 + c^2 \equiv 10 \pmod{16}$$

are inconsistent.

THEOREM 7. *If e, e_i $(i=1, 2, 3, 5)$, $e_4 = 2e_7$, s, s_i $(i=1, \cdots, 4)$, a, b, c have the same significance and satisfy the same restrictions as stated in Theorem 6, the set of congruences*

$$e_4^2 s_4^2 a^2 bc \equiv 4 \pmod 8, \qquad e_4^2 s_4^2 abc^2 \equiv 4 \pmod 8,$$
$$\tfrac{1}{4} e_4^2 s_4^2 a^2 bc \equiv h \pmod 8, \qquad \tfrac{1}{4} e_4^2 s_4^2 abc^2 \equiv k \pmod 8$$

is inconsistent when $h=3$ and $k=7$, $h=7$ and $k=3$, $h=5$ and $k=1$, or $h=1$ and $k=5$.

9. **Numerical examples.** Solutions* of equation (25) will now be given for a few values of n and p:

$$U^2 - acR^2 = (A^2 - abB^2)(F^2 - bcC^2), \tag{25}$$

where

$$a_1 = 2n - p = aca_0^2, \quad b_1 = -2n - p = bcb_0^2.$$

	n	p	a	b	c	U	R	A	B	F	C
(i)	−2	−1	−3	5	1	4	4	1	1	3	1
(ii)	1	−5	7	3	1	14	2	−7	1	3	1
(iii)	−4	−3	−5	11	1	10	6	1	1	4	1
(iv)	−4	−3	−5	11	1	0	80	−5	1	24	4
(v)	−2	1	−5	3	1	2	2	3	1	7	4
(vi)	−3	1	−7	5	1	12	60	2	−2	14	2
(vii)	1	−31	33	29	1	0	$30B$	$-33B$	arb. int.	6	3
(viii)	−4	3	−11	5	1	0	32	11	1	12	4
(ix)	−105	−294	3	2	7	315	35	30	5	28	7

Raymond J. Garver,† using different methods and different notation, obtained criteria for the solvability in integers of equation (25) when

$$U = 0, \quad ca_0b_0A = (p-2)B, \quad \text{and} \quad cb_0R \equiv 0 \pmod{2B}. \tag{40}$$

By imposing further restrictions on p and n, Garver treated the norm conditions for a division algebra and set up examples of division algebras of order sixteen satisfying all his conditions. All conditions (40) are satisfied by examples (iv) and (vii) above. In examples (ii) and (vi) all except the first of conditions (40) are satisfied. In example (viii) only the first and the third of conditions (40) are satisfied. In examples (i), (iii), (v), and (ix) (in which $a_0=2$, $b_0=6$), only the last of conditions (40) is satisfied. On taking $e=s=1$ we see that example (iv) satisfies all our previous conditions and, without modification, gives rise to a solution of equations (25) and (20), and of the original equation (12).

---

* When one solution of equation (25) is known for a given n and a given p, an infinite number of solutions are known; for, if $U=\lambda$, $R=\mu$, $A=\nu$, $B=\rho$, $F=\sigma$, $C=\tau$ is an integral solution, then also is $U=r_1r_2\lambda$, $R=r_1r_2\mu$, $A=r_1\nu$, $B=r_1\rho$, $F=r_2\sigma$, $C=r_2\tau$, where r_1 and r_2 are arbitrary integers.

† R. J. Garver, (I) *On Tschirnhausen Transformations*, (II) *Division Algebras of Order Sixteen*, Dissertation, Chicago, 1926.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

ON THE DEGREE OF APPROXIMATION TO AN ANALYTIC FUNCTION BY MEANS OF RATIONAL FUNCTIONS*

BY

J. L. WALSH

It is the purpose of this note to establish results for the approximation to an analytic function by means of rational functions analogous to the following theorem for such approximation by means of polynomials:

THEOREM I. *Let J be an arbitrary closed Jordan region of the z-plane, and let $w=\Phi(z)$ be a function which maps conformally the exterior of J onto the exterior of the unit circle in the w-plane so that the points at infinity correspond to each other. Let C_R denote the curve $|\Phi(z)|=R$, $R>1$, that is, the transform in the z-plane of the circle $|w|=R$.*

A necessary and sufficient condition that an arbitrary function $F(z)$ defined in J be regular-analytic in (the closed region) J is that there should exist polynomials $P_n(z)$ of degree n, $n=0, 1, 2, \cdots$, and numbers M, $R>1$, such that the inequality

$$|F(z)-P_n(z)| \leqq \frac{M}{R^n} \tag{1}$$

is valid for every z in J.

If the polynomials $P_n(z)$ are given so that (1) is satisfied, the sequence $P_n(z)$ converges everywhere interior to C_R and uniformly on any closed point set interior to C_R, and thus $F(z)$ is regular† throughout the interior of C_R.

If $F(z)$ is given regular in the closed interior of C_ρ, the polynomials $P_n(z)$ can be chosen so that (1) is satisfied for $R=\rho$ and for every z in J.

Theorem I is based on results and methods of S. Bernstein, M. Riesz, Faber, Fejér, and Szegö, but the formulation as just given is due to the present writer.‡

---

* Presented to the Society, September 7, 1928; received by the editors in April, 1928.

† Here and below we tacitly assume that if $F(z)$ is not originally supposed to be defined on the entire point set considered, then the definition in the new points is to be made by analytic extension—or what amounts to the same thing, by means of the convergent series of rational functions, polynomials in the present case.

‡ See Sitzungsberichte der Bayerischen Akademie der Wissenschaften, 1926, pp. 223–229, where detailed references to the literature are given.

A rational function of the form

$$r(z) = \frac{a_0 z^n + a_1 z^{n-1} + \cdots + a_n}{b_0 z^n + b_1 z^{n-1} + \cdots + b_n}$$

is said to be of degree n. We do not assume a_0 or b_0 different from zero, but do assume of course that the denominator does not vanish identically. If the poles of the function are all distinct, we may write $r(z)$ in the form

$$A + A_0 z + \frac{A_1}{z-\beta_1} + \frac{A_2}{z-\beta_2} + \cdots + \frac{A_n}{z-\beta_n},$$

where the A_i are not all different from zero; if the poles of the function are not distinct, we may write $r(z)$ in the form

$$A_0 + A_0' z + \cdots + A_0^{(k_0)} z^{k_0} + \frac{A_1'}{z-\beta_1} + \frac{A_1''}{(z-\beta_1)^2} + \cdots + \frac{A_1^{(k_1)}}{(z-\beta_1)^{k_1}} + \frac{A_2'}{z-\beta_2} + \cdots + \frac{A_m^{(k_m)}}{(z-\beta_m)^{k_m}}.$$

We proceed to the proof of

THEOREM II. *Let C be an arbitrary rectifiable Jordan curve of the z-plane in whose interior the origin lies, and let the functions $w=\Phi(z)$, $\phi(z)$ map respectively the exterior and interior of C onto the exterior and interior of the unit circle in the w-plane so that the point at infinity and origin of the one plane correspond to those of the other. Let C_R and D_R denote the curves $|\Phi(z)|=R$, $R>1$, and $|\phi(z)|=1/R$ respectively, that is, the transforms onto the z-plane of the circles $|w|=R$ and $|w|=1/R$.*

A necessary and sufficient condition that an arbitrary function $F(z)$ defined on C be regular-analytic on C, is that there should exist polynomials $p_n(z)$ in z and polynomials $q_n(z)$ in $1/z$ of degree n, $n=0, 1, 2, \cdots$, so that we have

$$|F(z) - p_n(z) - q_n(z)| \leqq M/R^n \tag{2}$$

for all z on C, where M and $R>1$ are constants independent of z.

If the polynomials $p_n(z)$ and $q_n(z)$ are given so that (2) is satisfied for all z on C, and if we modify (if necessary) these definitions so that $q_n(\infty)=0$, then the sequence $p_n(z)$ converges for all z interior to C_R and uniformly on any closed point set interior to C_R, and the sequence $q_n(z)$ converges for all z exterior to D_R and uniformly on any closed point set exterior to D_R; hence the function $F(z)$ is regular in the annular region bounded by C_R and D_R.

If $F(z)$ is given analytic in the closed region bounded by C_ρ and D_ρ, the polynomials $p_n(z)$ and $q_n(z)$ can be determined so that (2) is satisfied with $\rho=R$, for all z on C.

The proof is quite easy, by means of Theorem I. Let the polynomials $p_n(z)$ and $q_n(z)$ be given; assume, as we may do with no loss of generality, $q_n(\infty)=0$. We choose z on an arbitrary curve C' interior to C, and denote by d the minimum distance of C' from C. Then we have, for z on C',

$$(3)\quad \frac{1}{2\pi i}\int_C \frac{F(t)-p_n(t)-q_n(t)}{t-z}dt = f(z)-p_n(z), \text{ where } f(z)=\frac{1}{2\pi i}\int_C \frac{F(t)dt}{t-z}.$$

By (2) we now have

$$|f(z)-p_n(z)| \leqq \frac{Ml}{2\pi d R^n}$$

for z on C', where l is the length of C. It follows by Theorem I that the sequence $p_n(z)$ converges not merely on C', but within a curve C_R', the transform of the circle $|w|=R$ when the exterior of C' is mapped on the exterior of the unit circle of the w-plane so that the points at infinity correspond. The convergence of $p_n(z)$ is moreover uniform on any closed point set interior to C_R'. If now the curve C' is allowed monotonically to approach the curve C, the curve C_R' approaches monotonically the curve C_R.* Thus the sequence $p_n(z)$ converges interior to C_R and uniformly on any closed point set interior to C_R. It will be noticed that in particular the sequence $p_n(z)$ converges uniformly on C.

Consideration for z *exterior* to C of the integral used in (3) yields a similar result on the convergence of the polynomials $q_n(z)$, so the sequence $p_n(z)+q_n(z)$ converges between C_R and D_R, uniformly in any closed region between C_R and D_R, and $F(z)$ is regular-analytic between those two curves.

Conversely, let the function $F(z)$ be given analytic in the closed region bounded by C_R and D_R. Then $F(z)$ is also analytic in a larger closed annular region bounded by $C_{R'}$ and $D_{R'}$, $R'>R$. Cauchy's integral

$$f_1(z)=\frac{1}{2\pi i}\int_{C_{R'}}\frac{f(t)dt}{t-z}, \quad z \text{ interior to } C_{R'},$$

$$f_2(z)=\frac{1}{2\pi i}\int_{D_{R'}}\frac{f(t)dt}{t-z}, \quad z \text{ exterior to } D_{R'},$$

defines functions analytic in the closed regions interior to $C_{R'}$ and exterior to $D_{R'}$ respectively, and if the integrals are taken in the positive (counterclockwise) senses on those curves, we have

$$F(z)=f_1(z)-f_2(z).$$

* This follows easily from a comparison of the Green's functions for the two regions. Compare also Carathéodory, Mathematische Annalen, vol. 72 (1912), pp. 107–144, especially pp. 126–127.

Theorem I applied directly to $f_1(z)$, and to $f_2(z)$ after reciprocal transformation, yields now the polynomials $p_n(z)$ and $q_n(z)$ so that we have

$$(4)\qquad \begin{aligned} &|f_1(z) - p_n(z)| \leqq \frac{M_1}{R^n}, \quad z \text{ on or interior to } C, \\ &|f_2(z) + q_n(z)| \leqq \frac{M_2}{R^n}, \quad z \text{ on or interior to } C, \end{aligned}$$

which yields an inequality of form (2), for z on C.

Here it is not necessary to choose $q_n(\infty)=0$. That is, however, easy to accomplish if one cares to do it. For we have $f_2(\infty)=0$, so the new function $\mu(z)=z\,f_2(z)$ is analytic on and exterior to D_R. There exist polynomials $g_n(z)$ in $1/z$ such that

$$|zf_2(z) + g_n(z)| \leqq \frac{M_2'}{R^n}, \quad z \text{ on or exterior to } C.$$

We may now set $q_{n+1}(z)=g_n(z)/z$, a polynomial in $1/z$, and we have the equivalent of (4) because $|z|$ has on C a lower bound different from zero.

Theorem II is particularly simple when the curve C is itself the unit circle. In this case the expansion when $F(z)$ is given may be taken as a Fourier or Laurent series. This case has already been studied by de la Valleé Poussin* in much greater detail than we consider here. His results are of course more precise than Theorem II; those results can be used, in fact, to secure more precise results in the present situation, if C is analytic.

A less specific theorem than Theorem II, but more general in some respects, can be proved in a similar way. The reasoning we give includes in fact a proof of the sufficiency of the conditions of Theorems I and II.

THEOREM III. *Let C be an arbitrary rectifiable Jordan curve and let the function $f(z)$ be defined on C. If there exist rational functions $r_n(z)$ of degree $2n$, $n=0, 1, 2, \cdots$, so that we have for all sufficiently large n*

$$(5)\qquad |f(z) - r_n(z)| \leqq \frac{M}{R^n}, \quad R > 1, \text{ for all } z \text{ on } C,$$

and if the poles of the functions $r_n(z)-r_{n-1}(z)$ have no limit point on C, then the function $f(z)$ is meromorphic on C. If the functions $r_n(z)$ have no poles on C, then the function $f(z)$ is regular-analytic on C.

In determining whether a point is a limit point of the poles of the functions

* *Approximation des Fonctions*, Paris, 1919, Chap. VIII.

$r_n(z)-r_{n-1}(z)$, we count each pole according to its multiplicity in the totality of these rational functions.

We proceed with integration as in the proof of Theorem II. Under the present hypothesis the function $f(z)-r_n(z)$ is continuous on C for sufficiently large n, say for $n \geqq N$. We set

$$F(z) = f(z) - r_N(z), \quad r_n'(z) = r_n(z) - r_N(z), \quad n \geqq N,$$

and deduce

$$F(z) - r_n'(z) = f(z) - r_n(z), \quad n \geqq N.$$

Then we have for z on a curve C' interior to C,

$$|f_1(z) - p_n(z)| \leqq \frac{M'}{R^n}, \quad \text{where } f_1(z) = \frac{1}{2\pi i}\int_C \frac{F(t)dt}{t-z}, \tag{6}$$

$$p_n(z) = \frac{1}{2\pi i}\int_C \frac{r_n'(t)dt}{t-z}.$$

We have similarly for z on a curve C'' exterior to C,

$$|f_2(z) - q_n(z)| \leqq \frac{M''}{R^n},$$

where

$$f_2(z) = \frac{1}{2\pi i}\int_C \frac{F(t)dt}{t-z}, \quad q_n(z) = \frac{1}{2\pi i}\int_C \frac{r_n'(t)dt}{t-z},$$

the integrals taken in the positive sense on C. It follows of course that $q_n(\infty)=0$, and also that $r_n'(z)=p_n(z)-q_n(z)$ for z on C, if $p_n(z)$ and $q_n(z)$ are defined on C by analytic extension. For instance the integral defining $p_n(z)$ can be taken over C or over a larger Jordan curve in whose interior C lies, without changing the value of the integral.

The following reasoning is an extension of the reasoning used by Marcel Riesz* in connection with Bernstein's special case of Theorem I. Let us establish a lemma for use in studying the convergence of the sequence $p_n(z)$.

LEMMA. *Let Γ be an arbitrary Jordan curve of the z-plane, and denote by $w=\Phi(z)$, $z=\Psi(w)$, a function which maps the exterior of Γ onto the exterior of the unit circle γ in the w-plane so that the two points at infinity correspond. Let Γ_R denote the curve $|\Phi(z)|=R$, in the z-plane, for $R>1$. If $P(z)$ is a rational function of degree n whose poles lie exterior to ρ, $\rho>1$, and if we have*

$$|P(z)| \leqq L \text{ for all } z \text{ on } \Gamma,$$

then we have likewise

$$|P(z)| \leqq L\left(\frac{\rho R_1 - 1}{\rho - R_1}\right)^n, \text{ for all } z \text{ on } \Gamma_{R_1}, \ R_1 < \rho.$$

* Acta Mathematica, vol. 40 (1916), pp. 337–347.

The function $P[\Psi(w)]$ has at most n poles for $|w| \geqq 1$, and these all lie exterior to $|w| = \rho$. For convenience in exposition we shall suppose that there are precisely n poles $\alpha_1, \alpha_2, \cdots, \alpha_n$, not necessarily all distinct, and that none lies at infinity. If there are less than n poles, or if infinity is also a pole, there are only obvious modifications to be made in the discussion. In the latter case, for instance, we consider in the right-hand member of (7) the function found by taking the limit as one or more of the α_i become infinite. The function

$$\pi(w) = P[\Psi(w)]\frac{w - \alpha_1}{1 - \bar{\alpha}_1 w}\cdot\frac{w - \alpha_2}{1 - \bar{\alpha}_2 w}\cdots\frac{w - \alpha_n}{1 - \bar{\alpha}_n w} \tag{7}$$

is regular for $|w| > 1$ and continuous for $|w| \geqq 1$. The function

$$\frac{w - \alpha_i}{1 - \bar{\alpha}_i w}$$

has the absolute value unity for $|w| = 1$, so we have

$$|\pi(w)| \leqq L, \text{ for all } w \text{ on } \gamma\ ; \tag{8}$$

since (8) holds for w on γ it also holds throughout $|w| \geqq 1$. The transformation $\zeta = (w - \alpha_i)/(1 - \bar{\alpha}_i w)$ transforms $|w| = R_1$ into the circle $|(\zeta + \alpha_i)/(1 + \bar{\alpha}_i \zeta)| = R_1$, so we have

$$|\zeta| \geqq \frac{|\alpha_i| - R_1}{R_1|\alpha_i| - 1} \geqq \frac{\rho - R_1}{R_1\rho - 1}, \quad \text{for } |w| = R_1 < \rho.$$

Thus we find from (7) and (8),

$$|P[\Psi(w)]| \leqq L \prod_{i=1}^{n} \left|\frac{1 - \bar{\alpha}_i w}{w - \alpha_i}\right| \leqq L\left(\frac{R_1\rho - 1}{\rho - R_1}\right)^n$$

for $|w| = R_1 < \rho$, and the Lemma is established.

We return to the proof of Theorem III. From (6) we conclude

$$|f_1(z) - p_{n-1}(z)| \leqq \frac{M'}{R^{n-1}}, \quad |p_n(z) - p_{n-1}(z)| \leqq M'\frac{1 + R}{R^n},$$

for all z on C'. If the poles of $p_n(z) - p_{n-1}(z)$ lie exterior to C_ρ, $\rho > 1$ (and here we assume the fact not necessarily for all n but merely for all n sufficiently large), these poles lie likewise exterior to C_ρ', the transform of $|w| = \rho$ when the exterior of C' is mapped on the exterior of γ so that the points at infinity correspond to each other. There are at most $4n - 2$ of such poles, so we conclude from the Lemma

$$(9) \qquad |p_n(z) - p_{n-1}(z)| \leqq M' \frac{1+R}{R^n} \left(\frac{R_1\rho - 1}{\rho - R_1} \right)^{4n-2}$$

for z on the curve C'_{R_1}, $R_1 < \rho$. We use here the fact that $p_n(z)$ is defined by Cauchy's integral, as in (6), and the integral may be taken not merely over C but over any curve interior to C_ρ, so $p_n(z) - p_{n-1}(z)$ has no poles other than the poles of $r_n(z) - r_{n-1}(z)$ exterior to C_ρ. Thus the sequence $p_n(z)$ converges for z within the curve C'_ν, hence, as we see by allowing C' to approach C monotonically, for z within the curve C_ν, where $\nu = (1+\rho R^{1/4})/(\rho + R^{1/4})$, which is greater than unity. The convergence is uniform on any closed point set interior to C_ν.

If we know that $r_n(z)$ has at most n poles exterior to C, at least for sufficiently large n, we may replace the exponent $4n-2$ in (9) by $2n-1$, and take correspondingly $\nu = (1+\rho R^{1/2})/(\rho + R^{1/2})$. If these n poles of $r_n(z)$ exterior to C are coincident and do not change position with n, or more generally if we know that $p_n(z) - p_{n-1}(z)$ has at most n poles, we may replace the exponent in $4n-2$ in (9) by n, and thus set $\nu = (1+\rho R)/(\rho + R)$.

The convergence of the sequence $q_n(z)$ exterior to some curve $D_{\nu'}$ can be proved similarly. It follows that the original sequence $r_n(z)$ converges uniformly in an annular region in whose interior C lies, so $f(z)$ is meromorphic on C and is regular on C if the $r_n(z)$ have no poles on C.

We have really proved here a much more precise theorem than Theorem III. We do not state the most general theorem possible, but merely note that Theorem V (below) is also valid in the limiting case that the region C (of Theorem V) becomes a rectifiable Jordan curve.

Theorem III remains true if C is no longer a Jordan curve, but has a finite number of double points. For in Theorem III we have shown the original sequence of rational functions to converge uniformly not merely on C but in an annular region in whose interior C is contained. In the modified situation it is still true that the curve C lies interior to a region of uniform convergence of the original sequence $r_n(z)$.

Theorem III is not true if we omit the restriction that the poles of the $r_n(z)$ should have no limit point on C. We illustrate this fact by a simple example. Let C be an arbitrary Jordan curve. If we move a line not intersecting C parallel to itself until it just touches C, we have a point A of C on this line L, yet the entire interior of the curve C is in a half-plane bounded by L. Choose coördinate axes so that A is the origin and L is the axis of imaginaries, the interior of C to the left of L. The series

$$(10) \qquad \sum_{m=0}^{\infty} \frac{1}{4^m(z - 1/2^m)}$$

converges uniformly for all values of z on or to the left of the axis of imaginaries, for we have, for such points,

$$\left| z - \frac{1}{2^m} \right| \geqq \frac{1}{2^m}, \quad \frac{1}{|4^m(z - 1/2^m)|} \leqq \frac{1}{2^m}.$$

Thus (10) converges uniformly on C. Successive partial sums of series (10),

$$r_n(z) = \sum_{m=1}^{n} \frac{1}{4^m(z - 1/2^m)},$$

yield, for z on C, an inequality of type (5). The sum of series (10) is, however, not analytic at the origin and hence not regular everywhere on C. For as is well known, and indeed follows from Theorem III, series (10) represents a function regular over the entire plane, except that there are poles at the points $z = 1/2^m$ and hence there is an essential singularity at the origin.

Theorem I can be extended so that J is not a Jordan region, but a Jordan arc or indeed any closed bounded point set S whose complement with respect to the entire plane is simply connected. We can find a similar result for the approximation on S of functions by means of rational functions. The only necessary modification in the proof already given for Theorem III is to notice that although $\Psi(w)$ is not necessarily continuous for $|w| \geqq 1$, its absolute value is continuous. Let us state

THEOREM IV. *Let S be an arbitrary closed bounded point set which consists of more than a single point and whose complement with respect to the entire plane is simply connected. If there exist rational functions $r_n(z)$ of degree n, $n = 0, 1, \cdots$, such that we have*

$$|f(z) - r_n(z)| \leqq M/R^n, \quad R > 1,$$

for all z on S and for all sufficiently large n, and if the poles of the functions $r_n(z) - r_{n-1}(z)$ have no limit point on S, then $f(z)$ is meromorphic on S. If the functions $r_n(z)$ have no poles on S, then $f(z)$ is regular on S.

Let $w = \Phi(z)$ denote a function that maps the complement of S onto the exterior of the unit circle in the w-plane, so that the points at infinity correspond to each other. Let S_R denote the curve $|\Phi(z)| = R$, $R > 1$. If the poles of the functions $r_n(z) - r_{n-1}(z)$ have no limit point interior to S_ρ, then the sequence $r_n(z)$ converges interior to S_ν, where $\nu = (1 + \rho R^{1/2})/(\rho + R^{1/2})$, and the convergence is uniform on any closed point set interior to S_ν. Hence $f(z)$ is meromorphic in S_ν, and if $r_n(z)$ has no poles interior to S_ν, $f(z)$ is regular interior to S_ν. If the function $r_n(z) - r_{n-1}(z)$ has at most n poles, for n sufficiently large, we may set $\nu = (1 + \rho R)/(\rho + R)$, and in particular if the only poles of $r_n(z)$ are at infinity, we may set $\rho = \infty$ and $\nu = R$, so that the sequence $r_n(z)$ converges interior to S_R and $f(z)$ is regular interior to S_R.

Theorem IV yields almost at once

THEOREM V. *Let C be a closed region bounded by Jordan curves $C_0, C_1, \cdots, C_{k-1}$, such that no two of these curves have a common point and so that $C_1, C_2, \cdots, C_{k-1}$ lie interior to C_0. A necessary and sufficient condition that a function $f(z)$ be regular-analytic in the (closed) region C is that there should exist rational functions $r_n(z)$ of degree kn, $n=0, 1, 2, \cdots$, with no poles or limit point of poles on C, such that we have*

$$|f(z) - r_n(z)| \leqq M/R^n, \quad R > 1, \text{ for all } z \text{ on } C. \tag{11}$$

A necessary and sufficient condition that a function $f(z)$ be meromorphic in the region C is that there should exist rational functions $r_n(z)$ of degree kn, $n=0, 1, 2, \cdots$, with no limit point of poles of the functions $r_n(z)-r_{n-1}(z)$ on C, such that (11) obtains for sufficiently large n.

Denote by $z_1, z_2, \cdots, z_{k-1}$ arbitrary but fixed points interior to $C_1, C_2, \cdots, C_{k-1}$ respectively, and denote by $w=\phi_i(z)$ a function which maps conformally the interior of C_i on the interior of the unit circle on the w-plane so that $z=z_i$ corresponds to $w=0$. Denote by $w=\phi_0(z)$ a function which maps the exterior of C_0 on the exterior of the unit circle in the w-plane so that the points at infinity correspond. Let $\Gamma_R^{(0)}$ denote the curve $|\phi_0(z)|=R>1$, and $\Gamma_R^{(i)}$ the curve $|\phi_i(z)|=1/R<1$, $i=1, 2, \cdots, k-1$.

If the rational functions $r_n(z)$ are given so that (11) is satisfied for all z on C for n sufficiently large, and if the poles of the $r_n(z)-r_{n-1}(z)$ have no limit point exterior to $\Gamma_\rho^{(i)}$ and interior to $\Gamma_\rho^{(0)}$, then the sequence $r_n(z)$ converges in the region Γ exterior to $\Gamma_\nu^{(i)}$ and interior to $\Gamma_\nu^{(0)}$, where $\nu=(1+\rho R^{1/(2k)})/(\rho+R^{1/(2k)})$, and the convergence is uniform on any closed point set interior to this region. Hence $f(z)$ is meromorphic in Γ, and if $r_n(z)$ has no poles exterior to $\Gamma_\nu^{(i)}$ and interior to $\Gamma_\nu^{(0)}$, $f(z)$ is regular in Γ. If $r_n(z)$ has not more than n poles interior to $\Gamma_\rho^{(i)}$ and not more than n poles exterior to $\Gamma_\rho^{(0)}$, then we may set $\nu=(1+\rho R^{1/2})/(\rho+R^{1/2})$. If $r_n(z)-r_{n-1}(z)$ has not more than n poles interior to $\Gamma_\rho^{(i)}$ and not more than n poles exterior to $\Gamma_\rho^{(0)}$, then we may set $\nu=(1+\rho R)/(\rho+R)$. In particular if the only poles of $r_n(z)$ are at $z_1, z_2, \cdots, z_{k-1}$ and at infinity, not more than n in each point, we may set $\rho=\infty$, and $\nu=R$, so that the sequence $r_n(z)$ converges in the region Γ interior to $\Gamma_R^{(0)}$ and exterior to $\Gamma_R^{(i)}$; the function $f(z)$ is regular in Γ. In any case, for n sufficiently large the function $r_n(z)-r_{n-1}(z)$ can be expressed as the sum of k functions each of the form $p_n^{(i)}(z)-p_{n-1}^{(i)}(z)$, $i=0, 1, \cdots, k-1$, which has poles only interior to C_i (or for $i=0$, only exterior to C_0), and such that the sequence $p_n^{(i)}(z)$ converges everywhere exterior to $\Gamma_\nu^{(i)}$ (or for $i=0$, everywhere interior to $\Gamma_\nu^{(0)}$).

If $f(z)$ is given regular in the closed region Γ interior to $\Gamma_\nu^{(0)}$ and exterior to

$\Gamma_\nu^{(i)}$, the functions $r_n(z)$ may be found with poles only in the points z_1, $z_2, \cdots, z_{k-1}$ and at infinity so that (11) *is satisfied for all n for all z in C, with $R=\nu$. If $f(z)$ is given meromorphic in this closed region Γ, the function $r_n(z)$ may be found with poles only in the poles of $f(z)$ in this region, in $z_1, z_2, \cdots, z_{k-1}$ and at infinity, so that* (11) *is satisfied for sufficiently large n for z in C and $R=\nu$.*

In Theorem V it is not essential to suppose the bounding curves C_0, $C_1, \cdots, C_{k-1}$ rectifiable, for auxiliary rectifiable curves $C_0', C_1', \cdots, C_{k-1}'$ near to $C_0, C_1, \cdots, C_{k-1}$ respectively can be constructed interior to C. The theorem is true for the region bounded by these auxiliary curves. If these curves are allowed to approach $C_0, C_1, \cdots, C_{k-1}$ monotonically, we obtain Theorem V as stated. It is likewise possible to consider boundaries more general than Jordan curves, by this same limiting process.

Perhaps the necessity of the condition of Theorem V should be briefly discussed. If $f(z)$ is regular in Γ, Cauchy's integral for $f(z)$ for z in Γ automatically expresses $f(z)$ as the sum of k functions each analytic in Γ and which are respectively regular everywhere interior to $\Gamma_\nu^{(0)}$ and exterior to $\Gamma_\nu^{(i)}$. Theorem I yields the result desired, if we use the transformation $z'=1/(z-z_i)$ for the function regular exterior to $\Gamma_\nu^{(i)}$.

If $f(z)$ is meromorphic in the closed region Γ, we can write $f(z)=F(z)+r(z)$, where $F(z)$ is regular in Γ and $r(z)$ is rational, let us say of degree p. There exist rational functions $r_m'(z)$ of degree km such that we have

$$|F(z) - r_m'(z)| \leqq \frac{M'}{\nu^m} \text{ for all } m \text{ and for all } z \text{ on } C.$$

That is, we have

$$|f(z) - r_{m+p}(z)| \leqq \frac{M'}{\nu^m} \text{ where } r_{m+p}(z) = r_m'(z) + r(z).$$

We can consider $r_{m+p}(z)$ of degree $k(m+p)$, and write this last inequality

$$|f(z) - r_{m+p}(z)| \leqq \frac{M'}{\nu^m} = \frac{M}{\nu^{m+p}}, \quad \text{where } M = M'\nu^p.$$

This inequality holds for all z on C for sufficiently large index $n=m+p$ and is of type (11).

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

EXISTENCE AND OSCILLATION THEOREMS FOR NON-LINEAR DIFFERENTIAL SYSTEMS OF THE SECOND ORDER*

BY

WILLIAM M. WHYBURN

1. Introduction

Boundary value problems for second order linear differential systems have been investigated extensively by many authors of the past century. It is the object of the present paper to consider non-linear equations of the second order and to establish existence and oscillation theorems for these equations together with boundary conditions of the two-point type. The coefficients of the differential equations are taken as summable functions of the independent variable and all integrals in the paper are taken in the sense of Lebesgue. The quantities used are confined to the real domain.

2. Existence theorems

The following two theorems are special cases of general existence theorems that have recently been established by the author.†

THEOREM I. *Given the differential system*

$$dy_i/dx = \sum_{j=1}^{j=n} A_{ij}(x, y_1, \cdots, y_n)y_j \tag{1}$$

$$y_i(c) = a_i \qquad (i = 1, 2, \cdots, n), \tag{2}$$

where the functions A_{ij} are single-valued and continuous in $(y_1, \cdots, y_n)$ for each fixed x on X: $a \leqq x \leqq b$, summable (Lebesgue) in x on X for fixed $(y_1, \cdots, y_n)$, and bounded numerically for all x on X and for all $(y_1, \cdots, y_n)$ by a function $M(x)$ that is summable on X, and where $x = c$ is any fixed point of X and the quantities $a_1, \cdots, a_n$ are constants, then there exists a unique solution‡ of system (1), (2), provided that there exists a summable function $N(x)$ on X such that

---

* Presented to the Society, September 7, 1927; received by the editors in August, 1927.

† W. M. Whyburn, *On the fundamental existence theorems for differential systems*, to appear in Annals of Mathematics for December, 1928.

‡ By a solution of this system we mean a set of absolutely continuous functions on X that satisfies (2) and satisfies (1) *almost everywhere* on X.

$$|A_{ij}(x,y_1,\cdots,y_n)-A_{ij}(x,z_1,\cdots,z_n)|\leqq N(x)\sum_{r=1}^{r=n}|y_r-z_r| \qquad (i,j=1,\cdots,n)$$

for all $(y_1,\cdots,y_n)$ *and* $(z_1,\cdots,z_n)$ *and for every* x *on* X.

THEOREM II. *For the system*

$$dy_i/dx=\sum_{j=1}^{j=n}A_{ij}(x,y_1,\cdots,y_n;\lambda)y_j \qquad (i=1,\cdots,n), \tag{3}$$

$$y_i(c,\lambda)=a_i(\lambda) \tag{4}$$

where $a_1,\cdots,a_n$ *are continuous functions of* λ *on* $\bar{L}$: $\bar{L}_1\leqq\lambda\leqq\bar{L}_2$; *the functions* A_{ij} *are summable in* x *on* X *for each fixed* $(y_1,\cdots,y_n)$ *and each fixed* λ *on* $\bar{L}$, *continuous in* $(y_1,\cdots,y_n,\lambda)$ *for all* $(y_1,\cdots,y_n)$ *and all* λ *on* $\bar{L}$ *when* x *has a fixed value on* X, *bounded numerically by a function* $M(x)$ *that is summable on* X, *and finally there is a summable function* $N(x)$ *on* X *such that for all* λ *on* $\bar{L}$ *and for every* $(y_1,\cdots,y_n)$ *and* $(z_1,\cdots,z_n)$

$$|A_{ij}(x,y_1,\cdots,y_n;\lambda)-A_{ij}(x,z_1,\cdots,z_n;\lambda)|\leqq N(x)\sum_{r=1}^{r=n}|y_r-z_r| \qquad (i,j=1,\cdots,n),$$

there exists on $X\bar{L}$ *a unique set* $y_1(x,\lambda),\cdots,y_n(x,\lambda)$ *of continuous* functions such that for each fixed* λ *on* $\bar{L}$ *these functions are absolutely continuous in* x *on* X *and satisfy* (3) *almost everywhere on* X, *and such that the set of functions satisfies* (4) *everywhere on* $\bar{L}$.

COROLLARY. *If the hypotheses of Theorem* II *are met for every closed subinterval of the open* λ*-interval* L: $L_1<\lambda<L_2$, *then the conclusions of that theorem are valid for the domain* XL.

3. THE SECOND-ORDER SYSTEM

Consider the system

$$dy/dx=K(x,y,z;\lambda)z, \qquad dz/dx=G(x,y,z;\lambda)y, \tag{5}$$

where K and G satisfy the hypotheses of Theorem II and its corollary. Let $y(x,\lambda)$, $z(x,\lambda)$ denote the unique solution of (5) that satisfies the conditions $y(a,\lambda)=0$, $z(a,\lambda)=1$, for all values of λ on L.

To system (5) apply the transformation†

* Cf. Carathéodory, *Vorlesungen über reele Funktionen*, Berlin, 1918, p. 678.

† The use of this transformation in the study of boundary value problems was suggested to me by Professor H. J. Ettlinger.

$$y(x,\lambda) = u(x,\lambda)\sin v(x,\lambda),$$
$$z(x,\lambda) = u(x,\lambda)\cos v(x,\lambda).$$

This yields

$$(6)\qquad dv/dx = K\cos^2 v - G\sin^2 v,$$
$$du/dx = \frac{u[K+G]}{2}\sin 2v.$$

Carathéodory's existence theorem* yields at least one solution of (6) that satisfies $u(c,\lambda)=a_1$, $v(c,\lambda)=a_2$, where $x=c$ is a point of X and a_1 and a_2 are continuous functions of λ on L. Suppose that for some fixed λ on L there is more than one solution of (6) and these initial conditions. Let u_1, v_1 and u_2, v_2 be two such solutions and suppose for definiteness that they are not identical on C: $c\leqq x\leqq b$. Obviously the case where these solutions are identical on C but differ on $a\leqq x\leqq c$ can be obtained from the present case by modifying the notation slightly and inserting additional absolute value signs in the inequalities. Let H be a constant that is greater than any of the continuous functions 1, $|u_1|$, $|v_1|$, $|u_2|$, $|v_2|$, on X. Let $K_1=K(x, u_1\sin v_1, u_1\cos v_1)$, $G_1=G(x, u_1\sin v_1, u_1\cos v_1)$, $K_2=K(x, u_2\sin v_2, u_2\cos v_2)$, and $G_2=G(x, u_2\sin v_2, u_2\cos v_2)$. From (6) we have

$$v_1 - v_2 = \int_c^x [K_1\cos^2 v_1 - K_2\cos^2 v_2 - (G_1\sin^2 v_1 - G_2\sin^2 v_2)]dt,$$

$$\begin{aligned}|v_1 - v_2| &\leqq \int_c^x [|K_1 - K_2| + |G_1 - G_2| \\ &\qquad + (|K_2| + |G_2|)(|\sin^2 v_1 - \sin^2 v|)]dt \\ &\leqq 2\int_c^x (N[|u_1\sin v_1 - u_2\sin v_2| + |u_1\cos v_1 - u_2\cos v_2|] \\ &\qquad + 2M|\sin v_1 - \sin v_2|)dt \\ &< 4\int_c^x \left[N|u_1 - u_2| + 2(NH+M)\left|\sin\frac{(v_1-v_2)}{2}\right|\right]dt \\ &< 4\int_c^x H(NH+M)(|u_1 - u_2| + |v_1 - v_2|)dt,\end{aligned}$$

since $|\sin[(v_1-v_2)/2]|\leqq|(v_1-v_2)/2|$ and H is greater than unity. Similarly

$$|u_1 - u_2| < 4\int_c^x H(NH+M)(|u_1 - u_2| + |v_1 - v_2|)dt.$$

The foregoing inequalities are valid for each x on C.

* Loc. cit., p. 672 ff. This theorem does not apply to give uniqueness.

Let $v(x) = |u_1 - u_2| + |v_1 - v_2|$ and add the above inequaltities to get

$$(7) \qquad \begin{aligned} v(x) &< 8\int_c^x H[HN(t) + M(t)]v(t)dt, \\ v(x) &< 8\int_d^x H[HN(t) + M(t)]v(t)dt, \end{aligned}$$

where $x=d$ is the last point of C such that $v(x)\equiv 0$ on $c\leqq x\leqq d$. Let x_1 be any point of $d<x\leqq b$ and let x_2 be a point of $d\leqq x\leqq x_1$ at which the continuous function $v(x)$ takes on its maximum value for that interval. From (7)

$$v(x_2) < v(x_2)\int_d^{x_2} 8H[HN + M]dt \leqq v(x_2)\int_d^{x_1} 8H[HN + M]dt$$

and since $v(x_2)>0$, we may divide by it to get

$$\int_d^{x_1} 8H[HN(t) + M(t)]dt > 1.$$

This inequality must hold for every x_1 on $d<x_1\leqq b$ which evidently is impossible. Hence we have established the following theorem:

THEOREM III. *There exists a unique pair of functions $u(x, \lambda)$, $v(x, \lambda)$ each of which is continuous in (x, λ) on XL, absolutely continuous in x on X for each fixed λ on L, and such that the pair satisfies the conditions $u(c, \lambda)=a_1(\lambda)$, $v(c, \lambda)=a_2(\lambda)$ for every λ on L and satisfies system* (6) *almost everywhere on X for each fixed λ on L.*

THEOREM IV. *If $u(x, \lambda)$, $v(x, \lambda)$ is the solution of* (6) *such that $u(a, \lambda)\equiv 1$, $v(a, \lambda)\equiv 0$, then $u(x, \lambda)$ does not vanish on XL.*

Assume that $u(c, p)=0$. By Theorem III there exists a unique solution U, V of (6) on $a\leqq x\leqq c$ satisfying the conditions $U(c, p)=0$, $V(c, p)=v(c, p)$. The method of successive approximations yields a solution of this system for which $u(x, p)\equiv 0$ on $a\leqq x\leqq c$ and hence this solution must be identical with the one that we are considering. This, however, contradicts the fact that $u(a, p)=1$ and yields our theorem.

In the discussion which follows we will assume that $K(x, y, z; \lambda)$ and $-G(x, y, z; \lambda)$ are positive for all values of x and λ on XL and for all values of y and z. It follows immediately from (6) that for each fixed value of λ on L the function v is an increasing function of x on X. This gives a one-to-one continuous correspondence between the points of $v(a)\leqq v\leqq v(b)$ and the points of X. Furthermore, this correspondence preserves order so that if x_1 and x_2 are any two points of X such that $x_1<x_2$, then $v(x_1)<v(x_2)$. Since

$u(x, \lambda) > 0$ on XL the zeros of y coincide with the zeros of $\sin v$ and the zeros of z coincide with those of $\cos v$. If for a fixed λ on L, $(2m-1)\pi/2 \leqq v(b) \leqq m\pi$, where m is a positive integer, then y will have zeros for $v(x) = i\pi$ while z has zeros for $v(x) = (2i+1)\pi/2$, where $i = 1, 2, \cdots, m-1$. It follows immediately from the correspondence between v and x that the zeros of y and z separate each other on X.

CONDITION A. *There exists a value $\lambda = p$ on L such that if q is any fixed value of λ on $L_1 < \lambda \leqq p$ and M and $\overline{M}$, respectively, denote the upper bounds of K and $-G$ for this value of λ, then $M + \overline{M} < \pi/(b-a)$.*

CONDITION B. *There exists a subinterval D: $c \leqq x \leqq d$ of X such that if $m(\lambda)$ denotes the smaller of the lower bounds of K and $-G$ on D for each fixed λ, then $\lim_{\lambda \to L_1} m(\lambda) = +\infty$.*

Condition A insures that for every λ on $L_1 < \lambda \leqq p$, $v(b, \lambda) < \pi$ and consequently that $y(x, \lambda)$ does not vanish on $a < x \leqq b$ for fixed values of λ on $L_1 < \lambda \leqq p$. This follows since $v(b, \lambda) = \int_a (K \cos^2 v - G \sin^2 v)dt \leqq \int_a^b [M + \overline{M}]dt < (b-a)\pi/(b-a)$.

Condition B insures that $\lim_{\lambda \to L_2} v(b, \lambda) = +\infty$, since $v(b, \lambda) \geqq \int_c^d (K \cos^2 v - G \sin^2 v)dt > m(\lambda)\ (d-c)$ and $\lim_{\lambda \to L_1} m(\lambda) = +\infty$.

THEOREM V. *If K and $-G$ are positive on XL and for every y and z and if Conditions* A *and* B *are satisfied, then there exists an infinite collection, $k_1, k_2, \cdots$, of sets of characteristic numbers* for the system* (5), (8) *such that if p_i is any number in k_i, then $y(x, p_i)$ and $z(x, p_i)$ each have exactly i zeros on $a < x \leqq b$, and finally for each index i there exists an index j such that for every n greater than j, every number in k_n is greater than every number in k_i. System* (8) *is*

$$y(a, \lambda) = 0, \qquad y(b, \lambda) = 0. \tag{8}$$

As λ increases from a value between L_1 and p to L_2, $v(b, \lambda)$ increases from a value less than π to $+\infty$. If i is any positive integer, it follows from the continuity of $v(b, \lambda)$ in λ that for at least one value of λ on L, $v(b, \lambda) = i\pi$. Let k_i be the totality of values of λ on L for which $v(b, \lambda) = i\pi$ and let p_i be any value in k_i; then since $v(x, p_i)$ is an increasing, continuous function of x on X, it follows that for one and only one value of x on X, $v(x, p_i) = j\pi$, $j = 1, 2, \cdots, i$. Also, since $\lim_{\lambda \to L_2} v(b, \lambda) = +\infty$, we can determine q_i such that for every λ on $q_i \leqq \lambda < L_2$, $v(b, \lambda) > i\pi$ and hence no values of k_i can lie on $q_i \leqq \lambda < L_2$. We can also determine j such that $v(b, \lambda) < j\pi$ on $p \leqq \lambda \leqq q_i$

* By a characteristic number we mean a value of λ for which the system has a non-indentically vanishing solution.

and hence no values in k_j, k_{j+1}, $\cdots$ lie on $p \leqq \lambda \leqq q_i$. Hence for any $n > j$, every number in k_n is greater than every number in k_i. This completes the proof of Theorem V.

A consideration of the values of λ for which $v(b, \lambda) = (2i-1)\pi/2$ yields the following theorem:

THEOREM VI. *Under the hypotheses of Theorem* V *there exists an infinite collection of sets,* h_2, h_3, $\cdots$, *of characteristic numbers for the system* (5) $y(a, \lambda) = 0$, $z(b, \lambda) = 0$ *such that if* p_i *is any value in* h_i, $y(x, p_i)$ *and* $z(x, p_i)$ *each have exactly* i *zeros on* X *and such that for each* i *there exists an index* j *such that for every* n *greater than* j, *every number in* h_n *is greater than every number in* h_i.

By letting $y = u(x, \lambda) \cos v(x, \lambda)$, $z = u(x, \lambda) \sin v(x, \lambda)$ and repeating the foregoing discussion, we get existence and oscillation theorems of the above type for the system (5) together with the boundary conditions $z(a, \lambda) = 0$, $y(b, \lambda) = 0$ and with the boundary conditions $z(a, \lambda) = 0$, $z(b, \lambda) = 0$.

We close this section with an example of a non-linear system to which all of the foregoing theory applies. Consider the system

$$dy/dx = \lambda(\sin^2 y + 2)z/x^{2/3}, \qquad X: 1 \leqq x \leqq b,$$
$$dz/dx = -\lambda(2\cos^2 z + 3)y/x^{4/5}, \qquad L: 0 < \lambda < +\infty,$$

together with boundary conditions of any of the four types considered in this section. One easily sees that all of the hypotheses of the foregoing theorems are met by this system and hence Theorems II, III, IV, V, VI are valid. Many other examples can be constructed to illustrate the systems covered by this treatment.

4. THE LINEAR SYSTEM

The case where K and G do not depend on y and z is of especial interest in that it gives a new treatment of boundary value problems of the Sturmian type.* The linear case appears as a special case of all of the foregoing treatment and many of the hypotheses are automatically satisfied for this case. It is noted that in the present treatment all of the conditions on K and $-G$ are symmetric, which is to be expected when one notes the manner in which these coefficients enter into the system. Examples can be given of systems where conditions of the above type are highly desirable.†

* Cf. Bôcher, *Leçons sur les Méthodes de Sturm*, Paris, 1917, p. 66ff.

† Such an example arises in the minimum surface of revolution problem of the calculus of variations. Conditions for non-oscillation are desired for the system $y' = (1+\lambda^2)^{1/2} \cosh [x(1+\lambda^2)^{1/2} + \sinh^{-1}\lambda]z$, $z' = -(1+\lambda^2)^{1/2} \operatorname{sech} [x(1+\lambda^2)^{1/2} + \sinh^{-1}\lambda]y$. The usual conditions of K and G being positive cannot be satisfied for real values of x and λ while Condition A of the above theory can be applied to obtain the desired information.

THEOREM VII. *If for each fixed x_1 on $a<x\leqq b$, K and $-G$ are non-decreasing functions of λ on L and there exists a sub-interval $D(x_1)$ of $a<x\leqq x_1$ almost everywhere on which either δK or $-\delta G$ is positive, where δf is the increment in f due to a positive increment $\delta\lambda$ in λ, then $v(x_1, \lambda)$ is an increasing function of λ on L.*

Choose $\delta\lambda$ so that $\lambda+\delta\lambda$ lies on L. From (6)

$$d(\delta v)/dx = [\delta K \cos^2(v+\delta v) - \delta G \sin^2(v+\delta v)] - (K+G)\sin(2v+\delta v)\sin\delta v.$$

Using the continuity of v in (x, λ) and restricting the size of $\delta\lambda$, we may solve this equation to get

$$\delta v(x_1, \lambda) = e^{-\omega}\left\{\int_a^{x_1}[\delta K\cos^2(v+\delta v) - \delta G\sin^2(v+\delta v)]e^{\tau}dt\right\}$$

where

$$\omega = \int_a^{x_1}\{[K+G](\sin(2v+\delta v))(\sin\delta v)/\delta v\}dt,$$

$$\tau = \int_a^{t}\{[K+G](\sin(2v+\delta v))(\sin\delta v)/\delta v\}ds,$$

from which we conclude immediately that $\delta v(x_1, \lambda)>0$.

Under the hypotheses of Theorem VII there exists a one-to-one, order-preserving correspondence between the points of the v-interval and those of L for each fixed x on $a<x\leqq b$. This correspondence enables us to state that k_i and h_i of Theorems V and VI contain one and only one value of λ and that these values are ordered in the following fashion:

$$L_1 < k_1 < h_2 < k_2 < h_3 < \cdots < h_i < k_i < \cdots < L_2.$$

A transformation that was used by the author in another paper* can be used to show that $\phi(x, \lambda)=\alpha(x, \lambda)y-\beta(x, \lambda)z$ and $\psi(x, \lambda)=\gamma(x, \lambda)y-\delta(x, \lambda)z$, where α, β, δ, γ have simple properties as functions of x and λ, satisfy equations of the type of (5) (K and G are linear). Hence the foregoing treatment can be used to prove existence and oscillation theorems where the boundary conditions are $\phi(a, \lambda)=\phi(b, \lambda)=0$.

---

* These Transactions, vol. 30, pp. 630–640.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

ERRATA, VOLUME 24

O. C. HAZLETT, *A symbolic theory of formal modular covariants.*

Page 299, THEOREM III. The proof given is evidently incorrect, inasmuch as a polynomial in the differences of the ratios α_2/α_1, β_2/β_1, etc. is not necessarily a polynomial in the symbolic invariants $\alpha_1\beta_2-\alpha_2\beta_1$, $\alpha_1\gamma_2-\alpha_2\gamma_1$, etc. On the other hand, Corollary 1 of Theorem III follows at once from Lemma V. For any isobaric formal modular invariant M of a system of forms S which is of degree d_i in the coefficients of f_i and of weight w is identically congruent, modulo p, to $a_0^{d_1}$ multiplied by a linear combination, K, of products of the differences of the type $(\beta_2/\beta_1)-(\alpha_2/\alpha_1)$, where the α's and β's may be symbols arising from the same form or may be symbols arising from different forms; moreover, K is homogeneous in the symbols for each form, f_i, and such that each ratio α_2/α_1 for this form occurs in exactly d_i factors in each product, and, finally, K is symmetric, modulo p, in these symbols for each form f_i. Hence, by a well known result of classic invariant theory, any isobaric formal modular invariant of S is congruent, modulo p, to an algebraic invariant, A, of S, although A is not necessarily rational and integral. Then Theorem III follows from Corollary 1, since any algebraic invariant is expressible as a polynomial in the symbolic invariants $\alpha_1\beta_2-\alpha_2\beta_1$, etc.

OLIVE C. HAZLETT

ERRATA, VOLUME 30

C. R. ADAMS, *On the irregular cases of the linear ordinary difference equation.*

Page 526, equation (33). The minus sign between the two fractions should be a multiplication sign.

Page 527, equations (34). The superscript of all the γ's in these equations should be $(n-1)$.

Page 537. In the display six lines from the bottom of the page, $\gamma_1^{(n-2)}$ should be $\gamma_i^{(n-2)}$.